DOCUMENT RESUME

ED 186 700 CE 025 278

AUTHOR Pak, Moon J.; And Others
TITLE Development and Implementation of a Comprehensive Medical Review Program for FLEX. Final Report.
INSTITUTION Oakland Univ., Rochester, Mich.
SPONS AGENCY Health Resources Administration (DHEW/PHS), Bethesda, Md. Div. of Medicine.
PUB DATE 5 Apr 78
CONTRACT 231-75-0621
NOTE 77p.

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Adult Education; *Certification; Communication Skills; Cultural Education; Curriculum Development; *Diagnostic Teaching; *Foreign Medical Graduates; Individualized Instruction; *Medical Education; Medicine; Physicians; Postsecondary Education; Professional Continuing Education; Program Development; Reading Comprehension; Refresher Courses; Review (Reexamination); Vocational Education
IDENTIFIERS Federal Licensing Examination

ABSTRACT

A program at Oakland University, Rochester, Michigan, developed and adminstered a preparatory course for the Federal Licensing Examination (FLEX) for non-fully licensed physicians employed in the United States. Statistical data indicated that nearly 75% of the condidates for licensure in states employing FLEX were foreign medical graduates with a pass rate under 50% compared with a pass rate of 85% for U.S. medical graduates. The integrated basic and clinical science curriculum and English/cultural curriculum developed were presented during a three-month program. An elaborate diagnostic appraisal process identified participants' individual strengths and weaknesses prior to program participation. The pass rate of the fifty-nine participants who took the June, 1977, FLEX was 27.1%, approximately 22% below the overall anticipated national norm. Statistical analysis of the scores demonstrated significant improvement in clinical competence and clinical science with less improvement in basic sciences. A detailed comparison of individuals who passed with those who showed the least improvement on the FLEX suggested that certain characteristics such as age, number of years in the U.S., and number of years since graduation from medical school may influence examinees' likelihood of success on the FLEX even following participation in a comprehensive medical review program. (YLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED186700

FINAL REPORT

"Development and Implementation of a Comprehensive Medical Review Program for the FLEX"

(231-75-0621)

April 5, 1978

Submitted to: Ms. Jeanette Rayner
Project Officer
International Programs Staff
Division of Medicine
Bureau of Health Manpower
Health Resources Administration
Department of HEW
Federal Center Bldg. #2
4th Floor
3700 East-West Highway
Hyattsville, Maryland 20782

By: Moon J. Pak, M.D., Ph.D.
Program Director

Robert E. Church
Assistant Director

Arthur Lewandowski
Testing and Evaluation Consultant
Board of Coordinators

Arthur J. Griggs
Assistant Program Director

Dawn E. Bochel
Program Secretary

Center for Health Sciences
Oakland University
Rochester, Michigan 48063

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

CE025278

JUN 13 1980

TABLE OF CONTENTS

		Page
I.	Introduction	1
II.	Program Organization	2
III.	Program Development	3
	A. Student Identification	4
	B. Curriculum Development	6
	1. Medical Curriculum	6
	2. English Cultural Curriculum	7
	C. Diagnostic Appraisal of Students	8
	1. Appraisal of Medical Knowledge	8
	a. The Diagnostic Examination	9
	i. Creation of Item Pool	9
	ii. Examination Structure	11
	iii. Compilation Procedure	15
	b. Self-Appraisal of Medical Knowledge	16
	c. Pre-Course Information Regarding Appraisal of Medical Knowledge	17
	2. Appraisal of English Proficiency	17
	3. Information Feedback to Students	18
	D. Instructional Activity	19
	E. Student Affairs	21
	F. Evaluation Conference	22
IV.	Program Implementation	23
	A. Student Identification	23
	B. Curriculum Implementation	26
	1. Medical Curriculum	26
	2. English Curriculum	27

C. Diagnostic Appraisal of Students 28
1. Appraisal of Medical Knowledge 28
a. Previous FLEX Results 28
b. Letter of Reference 29
c. Self-Appraisal 30
d. Diagnostic Examination 30
2. Appraisal of English Proficiency 39
D. Instructional Activity 39
E. Student Affairs 41
V. Program Evaluation 44
VI. Discussion 59
VII. Summary 63
Bibliography 65

FIGURES

Figure 1: 12

The relative proportions of class time (*————*) and "sample examination" items in Hubbard (*------*) allocated to the six basic sciences.

Figure 2: 13

The relative proportions of class time (*————*) and "sample examination" items in Hubbard (*------*) allocated to the six clinical sciences.

Figure 3: 47

The trend of scores of persons taking FLEX more than once prior to the review course.

Figure 4: 53

Average basic science score for the "Best" and "Worst" groups in the Medical Review Program on their most recent pre-program FLEX attempt.

Figure 5: 53

Average clinical science scores for "Best" and "Worst" groups in the Medical Review Program on their most recent pre-program FLEX attempt.

Figure 6: 55

Average change in points for the "Best" and "Worst" groups on the clinical science and clinical competence sections of pre and post program FLEX examinations.

TABLES

Table 1: 14

The number of examination items selected for each science area of the Diagnostic Examination.

Table 2: 15

The relative distribution of question format types in Hubbard's "sample examination" and the Diagnostic Examination.

Table 3: 23

Summary of application process.

Table 4: 24

Personal characteristics of all program applicants.

Table 5: 25

Summary of participants' past examination and review program experience.

Table 6: 29

Mean scores of students in the review program on their most recent FLEX examination prior to entering the course.

Table 7: 31

Science areas checked off on the letter of reference forms as representing the participants' particular strengths (f = frequency of checks).

Table 8: 31

The frequency each peer rank category was checked for the candidates by their supervisors (number of candidates = 66).

Table 9: 31

Mean self-ratings for each science area by the students in the Medical Review Program (n = 45).

Table 10: 32

The frequency of using each step on the seven point rating scale for the 12 science areas of the Self Appraisal Form.

Table 11: 32

Speed of performance on the Diagnostic Examination administered at the beginning of the FLEX Review Program (n = 68).

Table 12: 32

Results of the split-half correlation of the Diagnostic Exam Part I and Part II as estimates of reliability.

Table 13: 34

The correlation matrix for the 12 science areas of the Diagnostic Examination and the pre-program FLEX scores of the students (n = 39).

Table 14: 35

The intercorrelation of scores of the Diagnostic Exam and the past FLEX for the 12 science areas for the students in the review program.

Table 15: 37

Diagnostic Examination results for the Medical Review Program participants.

Table 16: 38

The intercorrelation of four information sources pertaining to the appraisal of students' medical knowledge for each science area.

Table 17: 45

The comparison of means between FLEX scores of students before and after the Medical Review Program.

Table 18: 48

The comparison of means between the pre and post test administered to students in the FLEX review program.

Table 19: 49

Comparison of Pre and Post Diagnostic Examination performance as a function of question item format.

Table 20: 52

Comparison of characteristics between the "Best" (n = 13 passers) and the "Worst" (n = 15 lowest) students in the Medical Review Program.

Table 21: 52

Average pre-program FLEX scores for persons in the "Best" and "Worst" groups.

Table 22: 54

Average change in points for the "Best" and "Worst" groups on the June, 1977, FLEX in comparison to the pre-program FLEX scores.

Table 23: 57

Comparison of the Diagnostic Exam results for the "Best" and "Worst" groups.

Table 24: 57

Classification of "Best" and "Worst" students using the class means in each science area on the Diagnostic Exam as a cutting score.

Table 25: 58

Performance of the "Best" and "Worst" groups on the English Language Proficiency (ELP) and Aural Comprehension (AC) examinations in comparison to the entire class average.

APPENDIXES

Appendix A: Organizational Flow Chart

Appendix B: Application Form

Appendix C: Advertisement Flier, Letters, Ads

Appendix D: Admission Criteria Form

Appendix E: Integrated Curriculum

Appendix F: Diagnostic Examination

Appendix G: Self Appraisal Form

Appendix H: Letter of Reference Form

Appendix I: Michigan Test of English Language Proficiency

Appendix J: Daily Lecture Schedule

Appendix K: Recommended Reference List

Appendix L: Medical Schools Attended by Review Program Participants

Appendix M: Case History Handout

Appendix N: Self-Instructional Audio-Visual Materials in Diagnostic Radiology

Appendix O: Feedback Sheet on Diagnostic Examination

Appendix P: Handouts

Appendix Q: Newspaper Article on the Dismissal of State Employed Physicians in Illinois

Appendix R: Application Review System

I. INTRODUCTION

In July, 1975, the Bureau of Health Resources Development, Department of Health, Education and Welfare, awarded contract No. 231-75-0607 to Oakland University for the purpose of developing and administering a preparatory course for the Federation Licensing Examination (FLEX) for non-fully licensed physicians employed in the United States.

The FLEX is a three day medical competency examination used by 48 state medical practice boards for licensure to practice medicine. It is designed primarily for those physicians who choose not to follow the conventional pathway to state licensure, i.e., National Board Medical Examinations I, II and III, or for physicians who are ineligible to follow such a pathway. The conventional pathway is open only to graduates of United States or Canadian Medical Schools. Recent data indicate that close to 75% of the candidates for licensure in states employing the FLEX are now foreign medical graduates(FMGs) (1). FMGs who took the FLEX examination between June 1968 and December 1972 had a pass rate just under 50% compared with a pass rate of about 85% for United States medical graduates (USMGs)(1,2). Thus, the Federal government, as a function of its desire to upgrade the quality of Medical Care delivered in the United States, solicited proposals for the establishment of one or more preparatory courses for the FLEX examination.

The target population was those physicians whose employment would be restricted or eliminated were there a modification of regulations disallowing reimbursement for care delivered by those not meeting full licensure standards.

The objective of the proposal submitted by Oakland University was "to develop and implement a Comprehensive Medical Review Program geared at reviewing and expanding on the medical knowledge base of non-fully licensed physicians, subsequently improving their performance and passing rates on the Federation Licensing Examination." The activities identified as being necessary to accomplish this

objective included:

(1) development of a clinical science curriculum
(2) development of a basic science curriculum
(3) development of an English and cultural curriculum
(4) development of diagnostic and program evaluation examinations
(5) development of audio-visual teaching aids

The implementation of the program, as set forth in the proposal, would include:

(1) evaluation and review of each participant's knowledge base in clinical and basic sciences, along with their proficiency in English
(2) identification of specific areas of weakness for each individual, based on the results of the above "diagnostic process".
(3) establishment of an individually geared "prescriptive plan of study" utilizing the clinical, basic science, and English/ Cultural curricula
(4) lecture oriented instructional activity to fill the needs of the individual, supplemented with audio-visual material
(5) evaluation of the impact of this program on the participants

This final report reviews the developmental aspects of this program (where not provided in the quarterly reports I - VII) and furnishes a detailed account of program implementation. A discussion follows wherein the impact of this program is evaluated in relation to the actual performance on the June, 1977, FLEX by the program participants.

II. PROGRAM ORGANIZATION

Oakland University is a medium sized public institution located in Rochester, Michigan, twenty-eight miles north of Detroit. The university has a full-time faculty of 400 and a student population of approximately 10,000.

Predicated on the fact that Oakland University does not have a medical school or medical center, it was essential that the program framework be highly organized in order to maximize utilization of the universities resources as well as those of the adjacent medical community.

The organizational flow chart (Appendix A) indicates how personnel were assembled for the planning and implementation of the program. As can be seen

from the flow chart, the basic organizational unit of the program was the Board of Coordinators. The board consisted of five experts drawn from the basic and clinical sciences, English-cultural, audio-visual, and testing and evaluation services.

Each board member was charged with the responsibility of establishing a task force which would provide component information for the overall development of the program. Identification of task force members, curriculum advisors, and faculty was facilitated by experience obtained in providing a comprehensive medical review program for FMGs under a grant from the Michigan State Medical Society, Foreign Medical Committee in June, 1975.

Group leaders were identified according to the six basic science (anatomy, biochemistry, pathology, pharmacology, physiology, microbiology) and six clinical science (surgery, ob/gyn, pediatrics, medicine, ph/pm, psychiatry) disciplines represented on Day I and Day II of the FLEX examination, respectively.

The contract provided for a part-time Program Director and a full-time Assistant Director and Program Secretary.

Overhead was contained in the contract and consequently office space and access to university services was provided for program staff on campus.

III. PROGRAM DEVELOPMENT

Program development occurred in five distinct categories i.e., a) student identification, b) curriculum development, c) diagnostic appraisal, d) instructional activity, and e) student affairs.

Development activities culminated with a complete review of the program by national and local experts in an evaluation conference held just prior to program implementation.

A. Student Identification

The contract was quite explicit in that the program was to be geared for physicians serving on temporary, partial or institutional licenses or permits to practice medicine. These physicians were to be employed in county, municipal, state or other public or private hospitals, nursing homes, or in a designated area of need in a single state or group of two or more states and not in training programs approved by the council on Medical Education of the American Medical Association. Only bona fide refugee physicians in the United States on refugee visas were exempt from these requirements.

An Application Form was developed (Appendix E) which provided the admissions committee with the necessary information insuring the FLEX eligibility of each applicant.

An intensive advertising campaign was initiated approximately one year prior to the anticipated program starting date of January 31, 1977. Nationally, advertisements were placed in the Journal of the American Medical Association, the New England Journal of Medicine, and the bulletins of various FMG organizations. Advertisements were also carried in state and county medical publications. A mass mailing of informative letters and announcement fliers included many State Medical Practice Boards and Societies, Departments of Public Health, hospitals, and mental health institutions. Sample copies of advertisements, fliers, and descriptive letters are contained in Appendix C.

Procedures were established for the preliminary review of each application upon receipt. The medical school from which the applicant graduated was checked for its registration in the World Health Organization's Medical School Directory for institutional accreditation by the respective country. The applicants were then classified as follows, in order of decreasing acceptance priority:

Type A: Medical Underground

(1) Physician in U.S.
(2) Non-fully licensed (ECFMG certified) and/or FLEX registered
(3) Fully employed in stipulated activity

Type B: Potential Underground I

(1) Physician in U.S.
(2) Non-fully licensed (as above)
(3) In graduate training program (i.e., residency, etc.)

Type C: Potential Underground II

(1) Physician in U.S.
(2) Non-fully licensed (as above)
(3) Non-fully employed

Type D: State-Limited Licensure

(1) Physician in U.S.
(2) State Licensed
(3) Fully employed

Type E: Potential Underground III

(1) Physician not in U.S.
(2) Non-fully licensed (see above)

Beyond the criteria stipulated by the contract, the admissions committee chose to submit each application to a review based upon the additional information outlined on the Admission Criteria Form (Appendix D). Particular attention was paid to whether the applicant was ECFMG certified. Consideration was also given to the number of previous unsuccessful ECFMG or FLEX attempts and the achieved scores. The letters of recommendation were reviewed closely as they provided a great deal of information as to the motivation and academic potential of the students as well as any language or cultural difficulties to be anticipated. Finally, the students age and number of years after graduation from medical school were considered.

The committee determined the disposition of each application indicating whether the applicant had been accepted, rejected, or that the application was pending. All applicants were reviewed equally, with those judged eligible being

accepted on a first come-first serve basis.

Demographic data on the final student population will be presented in Section IV.

B. Curriculum Development

Curriculum development occurred simultaneously in two distinct categories:

1. Medical Curriculum

Without access to an actual FLEX examination, input was solicited from the Federation of State Medical Practice Boards, National Board of Medical Examiners, and physicians who had recently taken the FLEX, regarding the content and format of the exam. The purpose of this action was to develop a curriculum which reflected the knowledge base necessary to pass the examination.

The underlying premise in the development of the medical curriculum was that participants in the program had an appropriate four year undergraduate medical education and that the program would serve as a review and limited expansion of their knoweldge base. It was understood that a review course of short duration, regardless of its content and quality could not compensate for an inadequate undergraduate medical education.

After as thorough an investigation and analysis of the examination as possible, it was determined that the most effective method for presenting the program would be to integrate the basic anc clinical science subjects and present the didactic information via a systems approach to the human body. This approach appeared to be consistent with the trend in many undergraduate medical curricula.

In its final format, the Integrated Curriculum (see Appendix E) appears as a topical outline serving as a guideline for lecture preparation and orientation. It is composed of 12 separate units: introductory medicine, cardiovascular, respiratory, digestive, ob/gyn-genital, urinary, skeletal-muscular, hematopoietic, endocrine, central and peripheral nervous systems, special senses, and infectious

diseases.

The basic science disciplines represented in the Integrated Curriculum include: anatomy, biochemistry, microbiology, pathology, pharmacology and physiology. The clinical science disciplines include: medicine, ob/gyn., pathology, pediatrics, psychiatry and surgery.

2. English/Cultural Curriculum

The development of an English/Cultural curriculum was in direct response to the assumption that the majority of the participants in the program would be FMGs and that language and communication barriers contribute to lower pass rates on the FLEX.

The final curriculum was composed of four modules:

Module 1: Communication

The purpose of this module was to assist the doctor in discovering and exploiting his own most effective oral communication style and to increase awareness of the nature of the communication process so that he may consciously meet and solve the variety of communication problems he normally faces in his work. Two approaches would be utilized. First, each participant would be videotaped while conducting a mock interview with a patient. Secondly, a series of demonstration - discussions were planned covering:

(1) Interpersonal communication
(2) Cross-cultural communication
(3) Verbal and nonverbal elements
(4) Role and social relationships in communication
(5) Handling conflict
(6) Analysis of public speaking style

Module 2: Rapid Comprehension through Effective Reading

This segment of the curriculum was included to teach the necessary skills required to learn efficiently. Seven reading-learning skills would be covered in sequential presentations, with pre and post examinations utilized to gauge student development.

Module 3: Language - Spoken and Listening

Lectures on effective oral communication were planned with topics including:

(1) The sound system, stress, rhythm, and intonation patterns of American English
(2) Review and practice of common conversational sentence patterns most often misused by non-native speakers
(3) Increasing recognition and production of words through derivational affixes (Greek and Roman roots)
(4) Teaching medical terminology in lay language
(5) Increasing comprehension of regional and social dialects, particularly "Black English"

Module 4: Cultural

A series of six seminar-discussions were planned to introduce the foreign physician to American society and the American health care delivery system.

The six planned seminars included:

(1) The Medical Doctor and Federal Income Taxes
(2) Future Doctors
(3) Behing-the-Scenes Look at Hospital Administration
(4) Advantages of the Doctor's Involvement in Organized Medicine
(5) Race and Minority Relations
(6) Death and Dying

C. Diagnostic Appraisal of Students

The Medical Review Program was constructed such that it could be individually adjusted to meet each student's particular areas of deficiency, and still maintain its overall integrity as a review of medical information. This approach required a detailed analysis of each participant's level of competency in the medical sciences. In addition, the high percentage of FMGs anticipated in this program necessitated an evaluation of language skills relating to successful examination performance. As a result of this appraisal process, a prescriptive plan of study would be developed for each student.

1. Appraisal of Medical Knowledge

Since the primary goal of the program was to prepare physicians to pass the FLEX examination, the frame of reference for diagnostic appraisal was the manner

in which medical information is organized and expounded within an objective examination format, specifically FLEX.

Medical knowledge, as defined by the Federation of State Medical Boards includes the demonstration of proficiency in the basic sciences of anatomy, physiology, biochemistry, microbiology, pathology, and pharmacology; and the clinical sciences of medicine, obstetrics/gynecology, pediatrics, preventative medicine/public health, psychiatry, and surgery. Medical competency requires the judgement and application of this information to clinical situations and constitutes the patient management or clinical competency portion of the FLEX. Based on this schema, it was determined that the primary objective of the evaluation process was to identify each participant's relative strengths and weaknesses in the twelve basic and clinical sciences. The instruments used in this portion of the diagnostic appraisal are described below. Although program developers were familiar with the clinical competence portion of the FLEX, lack of expertise prevented the development of a valid instrument for evaluating judgement.

a. The Diagnostic Examination

The Diagnostic Examination (DE) was developed as one of four instruments for the appraisal of medical knowledge. The model followed for the construction of the DE was the sample FLEX examination as described by Hubbard (3). When necessary, changes were made to suit the specific needs of the review program.

i. Creation of Item Pool

Hubbard states that the initial phase of FLEX development involves the use of a committee of experts to formulate an outline of the intended test content. In addition, the relative importance of each category is decided so that a corresponding proportion of questions can be allocated to each. From this weighted outline, specific questions are formulated and returned to the committee for review.

In the Oakland University Program, the initial phase of test development began with the Integrated Curriculum. Since a substantial amount of effort and planning was directed at curriculum development, it was felt that it adequately reflected the full range of test content. Thus, it served as the content outline to assure an adequate coverage of information. Utilizing the basic science and clinical science task force group leaders, individuals were asked to identify approximately sixty (60) major areas or core concepts (CC) within their discipline. For each core concept, the group leader was to write two multiple choice questions dealing with the major area. In addition, the authors were asked to rate each item for its degree of difficulty. The rating scale consisted of four levels ranging from basic, fundamental information to highly sophisticated, specialized information.

The result of this procedure was that an item pool of approximately 1,500 questions was created. However, as Hubbard had cautioned, the quality, range, completeness, and usefulness of the items was extremely varied. The strategy used by different group leaders for compiling items seemed to correspond to some of these difficulties. If an individual made use of a sub-group of people to write items, the range of content was good; however the consistency of difficulty ratings, language use, and clarity varied. On the other hand, if a group leader took it upon himself to compile the entire list of core concepts and test items, the product was more consistent (either good or poor) but a restriction of content become obvious.

A second problem occurred which reflects the great amount of time required for this task. A number of individuals chose to make use of existing review questions drawn from a variety of commercial sources. This resulted in a large number of items which could not be clearly related to the curriculum outline and often lacked any rating of estimated difficulty.

ii. Examination Structure

The allotted time for the Diagnostic Examination was set at six hours, corresponding to one full day of the orientation session. Since the test was designed primarily to assess the baseline level of medical knowledge, the per question time allottment was estimated at 60 seconds rather than the 50-52 seconds per question allowed for the FLEX. This corresponded to a maximum of 360 items, with an equal distribution of 180 items each to the basic and clinical science sections. In order to be compatable with computer scoring programs, the final result consisted of 160 basic science items and 155 clinical science items.

The next element of examination structure dealt with the relative proportion of items that should be assigned to the various science areas. Two sources were examined to aid in this decision. Utilizing the Integrated Curriculum, the proportion of lecture time assigned to each of the twelve science areas was calculated. Secondly, the sample FLEX examination that is published in Hubbard (3) was analyzed with respect to its proportional test item distribution. This procedure, however, was subject to judgemental error, since many test items could justifiably fall into more than one science area, and the Hubbard exam was not necessarily meant to mirror the complete FLEX in all its proportions. The results of the above analysis are depicted in Figures 1 and 2. The two analysis appeared to be somewhat consistent.

For each science discipline, the average of the two results was used for the allocation of items on the Diagnostic Exam. Table 1 gives the resulting number of items assigned to each science discipline on the exam.

The final element of test structure deals with item format. As described in the FLEX Description of Examination pamphlet (4) and the Hubbard text (3), four item types appear on the FLEX (as with all other National Board examinations). Hubbard identifies these four types as: A) one best response; B) list matching; C) alternative matching; and K) multiple true-false.

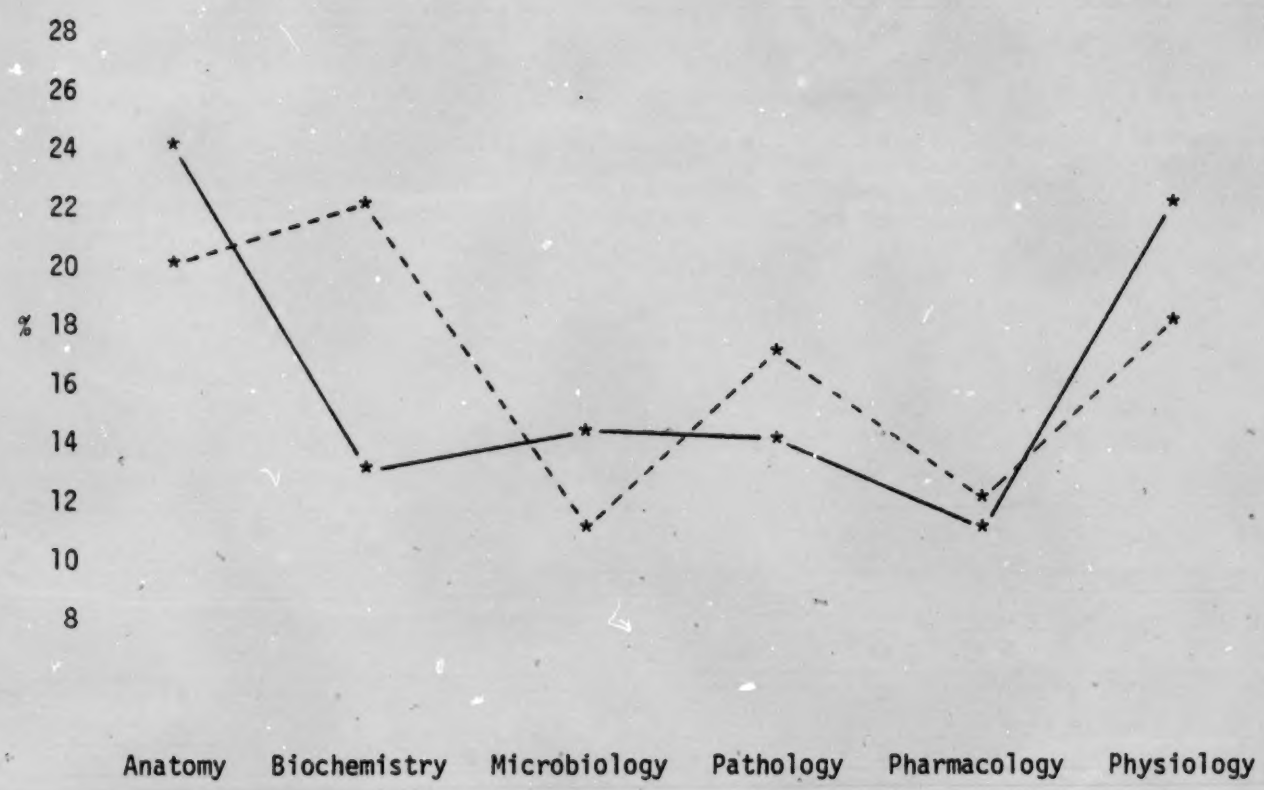


FIGURE 1: The relative proportions of class time (*————*) and "sample examination" items in Hubbard (*- - - - - -*) allocated to the six basic sciences.

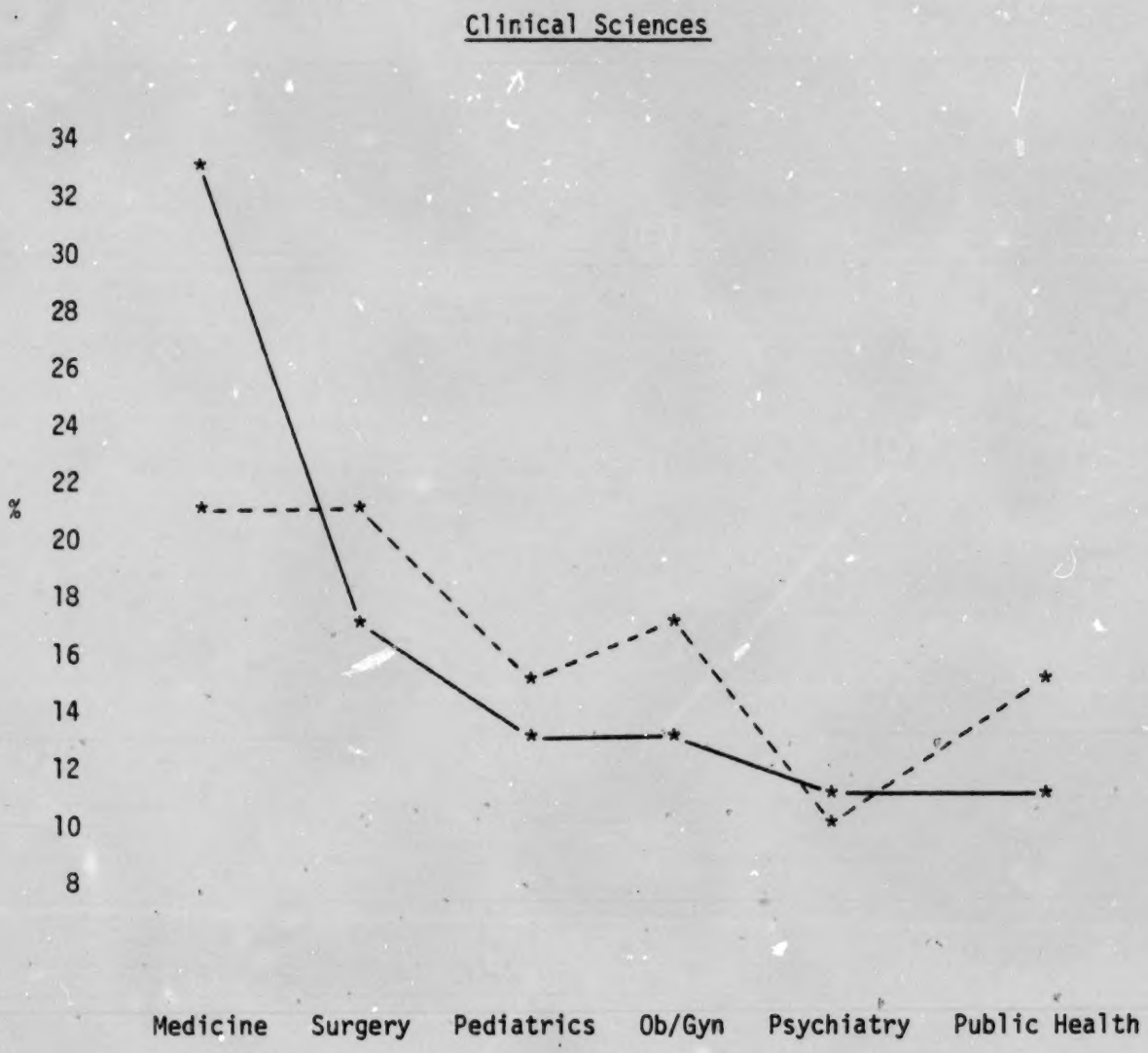


FIGURE 2: The relative proportions of class time (*————*) and "sample examination" items in Hubbard (*- - - - - - -*) allocated to the six clinical sciences.

	Number of Items		Number of Items
Basic Sciences	160	Clinical Sciences	155
Anatomy	34	Medicine	38
Physiology	32	Surgery	39
Biochemistry	28	Ob/Gyn	25
Microbiology	21	Pediatrics	21
Pathology	26	Ph/Pm	18
Pharmacology	19	Psychiatry	14

TABLE 1: The number of examination items selected for each science area of the Diagnostic Examination.

Again using the "sample examination" in Hubbard, the relative proportion of the test assigned to each item type was calculated, resulting in a prototype for the distribution of format in the Diagnostic Examination (Table 2). Constructing the test questions in the various formats proved to be a considerable problem. Almost all items contributed to the item pool by the group leaders were written in the one best response format. To convert an item into a different format, certain relationships were found which aided this procedure. First, an item written using the exclusion stem (i.e., all of the following EXCEPT...) could relatively easily be converted into a Type K question by dropping one of the five alternative answers. Secondly, as Hubbard had noted, certain material was better suited to different format types. For example microbiology questions lend themselves well to a matching format (Types B and C). The most difficult question type to contruct was the Type C Matching question. The difficulty arose from the fact that questions which ask one to differentiate between two alternatives, while the possibility that both and neither may be applicable, requires a sophisticated and precise knowledge about that information and in our procedure, the persons who contributed the information (i.e., group leaders writing core concepts and items) were

not the persons who dealt with the final test construction. An early attempt at re-cycling material back to the group leaders produced excessive delays and cumbersome procedures. The final conclusion reached was to collapse Type B and C formats together, thus allowing us to identify only three categories of question types.

Question Type	Percent in Model (Hubbard)	Actual Numbers of Questions in Diagnostic Examination	Actual Percent
A (One Best Response)	44%	207	66%
B (Matching)	29%	53	17%
C (Matching Both-Neither)	10%	11	4%
K (Multiple True-False)	17%	44	13%
	100%	315	100%

TABLE 2: The relative distribution of question format types in Hubbard's "sample examination" and the Diagnostic Examination.

iii. Compilation Procedure

Using the above structural requirements as a guideline, items were chosen from the item pool for the actual construction of the Diagnostic Exam. Care was excercised in choosing representative items from each of the major topical headings within each particular science. A grid system was formulated for this purpose, whereby question items from each of the 12 science areas were matched with the 12 units of the Integrated Curriculum.

The criterion of question difficulty was applied at this point in item identification. As described earlier, contributors of questions were

directed to rate each item for difficulty level using a four point scale. The decision was made prior to item selection that the majority of questions should be of moderate difficulty (Level II and III) with lesser amounts devoted to fundamental (Level I) and to highly specialized information (Level IV). A target proportion arrived at was 10 percent Level I, 80 percent Level II and III, and 10 percent Level IV.

Using the science of surgery as an example, the resulting Diagnostic Examination item profile was as follows: 1) 39 questions in total; 2) 17 one best response, 15 matching,and 7 multiple true-false questions; 3) 3-4 questions drawn from each of the twelve units of the Integrated Curriculum; and 4) 4 Level I, 31 Level II and III, and 4 Level IV questions.

The selected items for the Diagnostic Examination were carefully analyzed for duplication of information, consistency of language, and clarity of statements. A draft was then constructed by randomly mixing questions for order, with format type being kept segregated. Basic sciences and clinical sciences were dealt with as separate booklets. Following review and approval of the working drafts by representatives from each science area, final editing and duplication was performed for use of the examinations in the FLEX Review Program.

Instructions and procedures for the administration of these examinations were as follows: 1) each test (basic and clinical) is to be given in one sitting; 2) machine scoreable answer sheets with explanation are to be used; 3) a statement of caution regarding the test as a prototype for experimental use is made; 4) no time limit (within reason) is set; and 5) results will be reported back to the students in the form of raw scores and standard scores derived from the class as a group for each of the 12 science areas.

A copy of the Diagnostic Examination is contained with Appendix F.

b. Self-Appraisal of Medical Knowledge

A self-appraisal by each participant of his or her proficiency in

each of the basic and clinical sciences was felt to be an important consideration for prescriptive planning. To obtain this information, two sources were utilized. First, on the Application Form, participants were asked to identify the science areas which are the least difficult for them. Secondly, a Self-Appraisal Form (Appendix G) was constructed and mailed to each participant for completion prior to the start of the program. The topics of medical knowledge used in this form were taken directly from the FLEX descriptive pamphlet (4). The respondant was asked to rate himself with regard to each item using a seven point scale. The scale combines the two dimensions of knowledge and experience and uses ones peers as a reference point (score 4). The combination of dimensions was chosen in order to encompass both the young, newly graduated medical student and the older practicing clinician. For each science area, a mean rating could be calculated in order to have comparable scores to rank the respondant's relative strengths and weaknesses in relationahip to the group.

c. Pre-Course Information Regarding Appraisal of Medical Knowledge

Two additional sources of information regarding each participant's level of medical knowledge were the letter of recommendation furnished by a supervisor and, if available, past FLEX scores obtained from the Federation of State Medical Boards (FSMB). On the Reference Letter Form (Appendix H), the supervisor was asked to identify the applicant's areas of particular strength, again using the 12 sciences by title. The past FLEX results were obtained from the records of the FSMB through the cooperation of Dr. Frederick Merchant. This information was secured with the knowledge that it would be used for research purposes and evaluation and would be handled with strictest confidence.

2. Appraisal of English Proficiency

Diagnostic appraisal of English proficiency was conducted for the purpose of assessing each participant's language and communication skills. Providing English instruction to those participants demonstrating language deficiency

was visualized as an important supplement to the review of medical knowledge in improving the success rate on FLEX. In addition, this instruction would be beneficial to the physicians within the bounds of their workload during this program and during their everday affairs.

The instruments chosen for appraisal of English proficiency were the Michigan Test of English Language Proficiency and the Michigan Test of Aural Comprehension as offered by the English Language Institute of the University of Michigan (Appendix I). The standards used in evaluating student scores were those recommended by the English Language Institute for graduate students in the sciences.

As a result of their performance on the examinations, students were classified as follows:

(1) Passed the examination with no recommended English language tutoring
(2) Highly recommended for English language tutoring
(3) In definite need of English language tutoring

Students in categories (2) and (3) were strongly urged to participate in one or more of the English modules, particularly Module 3: Language - Spoken and Listening. The lectures in this module would be reinforced by continuous one-to-one tutoring for each participant by graduate students majoring in the teaching of English as a foreign language for the duration of the program.

3. Information Feedback to Students

As a consequence of the diagnostic appraisal of medical knowledge, four parallel lists, each ranking the 12 sciences in order of relative strength vs. weakness, were presented to each participant. The lists included: 1) Diagnostic Examination, 2) Self-Appraisal, 3) past FLEX, and 4) Supervisor Rating. By ordering the relative position of the sciences under these various categories, each student was able to identify areas which consistently appear strong or weak, identify unstable areas which appear to fluctuate, and identify possible sources of the instability (e.g., how I see myself vs. how my supervisor sees me). From this information, the emphasis in study and options for review were planned.

Since this feedback procedure was viewed as the initial model for program planning, individual half-hour sessions with each student were arranged in order to more closely monitor the information use and each student's reactions.

The results of the diagnostic appraisal of English proficiency were forwarded directly to the students in the form of the classifications as described above. The clearcut subdivision of the student population into the three categories of English participation required no further feedback.

D. Instructional Activity

Development of the medical and English curricula and total program planning occurred with the understanding that the contract budget allowed a 13 week, 65 day program.

After considerable discussion it was agreed that didactic presentations should not exceed 6 hours per day, resulting in a maximum of 390 lecture hours available to cover both the medical and English curricula.

Guided by the FLEX content and format, the Basic and Clinical Science and English/Cultural Task Forces apportioned these available lecture hours to the units of the Integrated Curriculum and the English/Cultural modules. A Daily Lecture Schedule (Appendix J) resulted.

It was apparent during the early stages of program development that the Integrated Curriculum, combined with the Daily Lecture Schedule, would be advantageous both in the administration of the program and in shaping proper study habits. Information provided on the Daily Lecture Schedule included the section of the Integrated Curriculum to be covered by any particular lecturer. By knowing beforehand the topics to be covered during a lecture, the students were encouraged to read and prepare for each lecture in advance.

This concept was carried one step further by assimilating these materials into a Pre-Program Study Package which was sent to each participant approximately

three months prior to the start of the program. Access to this material well in advance of the program starting date allowed each student to begin a systematic review of the material at home, identify areas of strength and weakness, and formulate specific questions to present to the faculty.

Each task force group leader assumed the responsibility of identifying competent faculty to present the material under the pervue of that task force. Eighty lecturers were identified and informed well ahead of time of the specific content of their lecture and how that particular lecture integrated with previous and following presentations and the program as a whole.

Faculty were instructed that in as much as the desire would exist to cover as much information as possible, time and program design only allowed for a brief review and an update of current theories, practices, and procedures. The actual methodology for each lecture presentation was left to the discretion of the faculty. However, faculty were encouraged to either develop content specific audio-visual materials or to use materials currently available commercially. The Audio-Visual Task Force stood ready to assist in this regard.

Faculty were also reminded that the primary objective of the physician participants would be to pass a multiple choice, objective type examination and that an opportunity should exist to allow the students to ask questions following each presentation or to challenge the students with content specific questions in a format consistent with the different FLEX item types.

To accomplish this objective, sample FLEX questions were made available to each instructor who was not familiar with the exam.

In addition, instructors were encouraged to submit questions relative to their presentation (using the FLEX item type formats) which would be compiled at the end of each system into a series of "Mock Examinations." The objective of these "Mock Examinations" was primarily to provide practice to the students in the art of taking an objective multiple choice examination compatable to the FLEX, since

it had been suggested that the format of the examination itself served as an additional barrier to achieving passing scores. Secondarily, these exams would serve as a motivating force, since each student would be comparing his progress and knowledge of a particular system with his immediate peers.

E. Student Affairs

The previous sections of this report have dealt with efforts to coordinate and develop a program of material designed explicity to prepare the participants for successful performance on the FLEX.

Concommitant with the development of the curricula and program methodology to accomplish the main objective, a good deal of effort and planning was devoted to creating an enriched living environment that would be conducive to learning.

It was the concensus of the Board of Coordinators that program participants should be encouraged to live in dormitory rooms provided by the university, as this would provide each student with an interesting cultural experience. In addition, many participants with vast experiential backgrounds and specialty training could serve as an available educational resource.

Students were paired two to a room. A conscious effort was made to separate speakers of the same language, which was intended to prohibit the students from reverting to their native tongue and force them to speak English. In addition, considerable attention was paid to combining roomates of compatable nationalities. This was accomplished with the guidance of the International Institute of Detroit.

Realizing that many of the participants would not possess current textbooks, a Recommended Reference List (Appendix K) was included in the Pre-Program Study Package and the university bookstore was prepared to handle text requests. It was also realized that some of the participants could not afford current texts and

consequently, at least two copies of each recommended reference were purchased and a private library was established in the dormitory.

Anticipating the great deal of anxiety and stress that would be experienced by these physicians, a clinical psychologist was retained to counsel the students should any difficulties arise.

Program staff assumed the responsibilities for coordinating the day to day personal requirements of the students, i.e., check cashing, transportation, audio-visual duplication, etc.

Though Oakland University does not have a medical school or medical center, the benefits to the students of being able to live comfortably on campus, walk to classes and have access to all the university's resources, ie., library, theatre, sports building, etc., contributed greatly to the attributes of the program.

F. Evaluation Conference

The contract stipulated that program activities be divided into two phases; the Developmental Phase and the Implementation Phase.

The Developmental Phase spanned a period of approximately 18 months and was for the design and development of the program. After 12 months of developmental activity, an Evaluation Conference was held where experts in medical education, educational psychology, English, sociology, testing, etc., were invited to review the program format and content. It was intended that the remaining six months of the Developmental Phase be utilized to incorporate the revisions and modifications to the program recommended at the conference.

The program itinerary consisted of a dinner reception with a keynote address being presented by Dr. Frederick T. Merchant, Chairman, FLEX Test Committee.

The evaluation conference proved to be very beneficial and resulted in ten constructive modifications to the original program design (see Quarterly Report V, p. 3).

IV. PROGRAM IMPLEMENTATION

A. Student Identification

Table 3 provides a survey of the results of the program advertising campaign and summarizes the student application process.

Category	Number
Application Requests	454
Applications Returned	157 (100%)
Withdrawals - no further response . . .	43 (27%)
Rejections	25 (16%)
Accepted Alternates	9 (6%)
Accepted Applicants Committed to Program	80 (51%)

TABLE 3: Summary of application process

As can be seen from Table 3, the final number of applicants registered to attend the program amounted to 51% of the final application pool.

The principle factors which accounted for individuals being judged as not acceptable for the program included:

(1) previously or currently licensed in a state to practice medicine
(2) not a United States citizen or not meeting immigration status as stipulated by the contract
(3) inappropriate or confusing application which could not be clarified or validated
(4) over 62 years of age

Of the 80 students accepted, 74 appeared for registration in February, 1977. Two individuals subsequently withdrew, resulting in 72 physicians completing the program.

Table 4 shows a comparison of the personal characteristics of all program applicants. The mean age of the class was 46.7 years, which was somewhat older than anticipated. Two-thirds were men and one-third women. The majority were married (67%) and had been in the United States an average of 14 years.

Variable		Class	Alternates	Withdrawn No Response	Rejection
Sex:	Male	66%	66%	82%	71%
	Female	34%	34%	18%	29%
Marital Status:	Married	67%			
	Single	33%			
Age Mean:		47.6	40.5	46.0	45.4
Age Range:		26-60	31-62	29-63	27-75
Current Residence:	Midwest	79%	78%	57%	48%
(U.S. Region)	East	5%	11%	18%	14%
	South	8%	0%	14%	10%
	West	8%	11%	11%	19%
	Non U.S.	0%	0%	0%	9%
Years in U.S.	Mean	14.0	14.6	12.0	14.25
	Range	1-53	1-54	1-49	0-54

Geographic Area of Origin:

Europe	16%
Mid East	8%
E. Asian	45%
S. America	10%
C. America	15%
N. America	5%

TABLE 4: Personal characteristics of all program applicants

The greatest proportion of the class resided in the Midwest (79%) with the state of Illinois accounting for over half of the participants (58%). The class represented a variety of nationalities with East Asian people accounting for 45% of this group.

Forty-five different medical schools were represented by the class (Appendix L). The average number of years since graduation from medical school was 18.6 years and ranged from 2 to 37. Eight-nine percent of the class reported having completed an internship and 71% had been in a residency training program. Seventy-four percent of the class were currently practicing medicine, 5% were in training positions, 13% were engaged in para-physician positions (i.e., physicians' assistant, surgical technician, orderly, etc.), and 8% were either unemployed or working in a non-medical occupation.

For the majority of the class, this program was the first formal review course they had taken. Only three persons in the class (all recent Vietnamese refugees) had no prior experience with ECFMG, FLEX, or previous review programs. Of the 49 persons who had previously taken the FLEX, a breakdown of the data for their most recent attempt shows that about one half of them had not attempted the examination during the past year, referring to the three scheduled periods after December, 1975. This information is summarized in Table 5.

Number of Previous Examinations/Review Programs (n=73)

	0	1	2	3	4	5 or more
ECFMG	12	42	7	6	5	1
FLEX	24	23	15	8	1	2
Review Programs . . .	49	18	3	3	0	0

Date of Most Recent FLEX Attempt (n=49)

12/76	6/76	12/75	6/75	12/74	6/74	1973	1972 or earlier
1	15	8	9	5	1	5	5

TABLE 5: Summary of participants' past examination and review program experience.

B. Curriculum Implementation

1. Medical Curriculum

The Integrated Curriculum proved to be an effective guideline for the presentation of didactic basic and clinical science information and for student self-study in preparation for Days I and II of the FLEX. However, it became apparent early in program development that the curriculum itself was inadequate for dealing with Day III, clinical judgement. Student feedback reinforced the fact that a definite need existed to supplement each system review with case history presentations in a format consistent with Day III of the FLEX.

Another shortcoming of the Integrated Curriculum was the lack of a built in opportunity to review and build on fundamental skills wuch as roentgenography and electrocardiography.

Efforts to compensate for shortcomings in the curriculum included the additional scheduling of a few evening and Saturday morning sessions in which a clinician reviewed relevant patient case histories corresponding to the week's system review. An excellent example prepared by a Nephrologist from Henry Ford Hsopital is presented in Appendix M.

As eluded to earlier in this report, a means of evaluating clinical judgment was beyond the scope of this program. Likewise, little could be done in regard to preparing valid patient management problems for class review. The basis for this inadequacy stems from the sparsity of information concerning the Day III portion of the FLEX.

One commercial source of patient management case problems was identified (5) and made available for student use. In addition, a series of self-instructional, audio-visual materials dealing with diagnostic radiology was purchased and made available to the students (Appendix N).

2. English Curriculum

The English curriculum was implemented as originally designed, with the exception of Module 1: Communication. It was the intention of the English/Cultural faculty to present this module during the three day orientation period and it would be accomplished by video-taping each participant in a simulated doctor-patient relationship for the purpose of evaluating communication effectiveness. This component of the program was eventually cancelled due to a strong feeling on the students' part of a need to set priorities on the use of their time, and a genuine anxiety over their patient management competence being evaluated and permanently recorded. In addition, there was a mistaken belief among the participants that this program was under the quasi-official auspices of the Federation of State Medical Boards, due to this organization's input and advise in program development and this evaluation would in some part effect the scoring of their examination.

Student response to the English/Cultural curriculum was quite mixed. It was generally agreed that the one-on-one tutorial sessions offered in conjunction with Module 3: Language, were most beneficial and that Module 2: Rapid Comprehension Through Effective Reading was helpful when presented early in the program to reinforce the students' study habits.

Most of the negative criticism of the English/Cultural curriculum occurred in relationship to Module 4: The Cultural Seminar discussions. Participation in these sessions was very poor and occurred as a result of a faulty assumption in program planning that the majority of the participants would be relatively young, recent immigrants to the United States. As the data in Table 4 clearly indicates, however, the median age was 47.6 years and the average student had been in the U.S. approximately 14 years. Consequently, the physician participants saw very little relevance of this module to their impending success on the FLEX and, as time was at a premium, most chose not to attend.

C. Diagnostic Appraisal of Students

1. Appraisal of Medical Knowledge

The appraisal of each student's medical knowledge was conducted with the information derived from the four sources discussed above: 1) previous FLEX scores, 2) Letter of Reference, 3) Self-Appraisal Form, and 4) the Diagnostic Examination.

The 30 minute meeting conducted between each student and the clinical coordinator of testing and evaluation was spent in an analysis and discussion of the diagnostic information available. Each information source ranked the 12 sciences in order of relative strength vs. weakness. When marked inconsistencies in ranking occurred between the four sources, discussion centered around the possible explanations (e.g., time spans, changes in training or work, changes in approaches to reviewing the material).

During this period of individual interaction between student and program coordinator, a great deal of significant, personal issues were raised and discussed. These issues will be elaborated upon in the section of this report dealing with student affairs.

The tremendous amount of information derived from the diagnostic evaluation process not only functioned as the focal point of individual student evaluation, but just as importantly, provided valuable data about the body of students as a whole. The correlations drawn between the information compiled and the overall success of the group on the FLEX are obvious. Therefore, discussion of each diagnostic procedure follows, providing results of the group of students as a whole.

a. Previous FLEX Results

Through the cooperation of the Federation of State Medical Boards, past FLEX scores were obtained for 44 of the 74 persons initially enrolled in the program. One of these physicians had taken only Day II of the exam previously. Therefore, information on Day I and Day III was complete for 43 students.

Analysis of this group as a whole indicated that the mean FLEX weighted average was 68.0 with the majority ranging between 64.5 to 71.4. A breakdown of scores for the science areas (Table 6) suggests that, in general, the basic sciences of anatomy, physiology and biochemistry, and the clinical sciences of psychiatry, public health, and surgery were the relative areas of weakness in the group. As was expected, the basic sciences were the lowest scores while clinical sciences were intermediate to the clinical competence (Day III) results. A notable point within this data pertains to psychiatry. Over half of the students in the program had been clinically employed in psychiatric settings. The past FLEX performance in this science discipline has a concomitant increased variability of scores when compared to the other sub-scores.

	(n=43)			(n=44)	
	Mean	S.D.		Mean	S.D.
Anatomy	59.8	5.52	Medicine	69.4	4.95
Physiology	58.6	5.39	Surgery	68.9	5.32
Biochemistry	57.3	5.28	Ob/Gyn	69.6	4.83
Microbiology	61.9	4.52	Ph/Pm	66.2	4.48
Pathology	66.1	5.42	Pediatrics	69.6	5.57
Pharmacology	65.6	5.62	Psychiatry	65.2	8.91
Basic Science Average	61.6	4.07	Clinical Science Average	68.2	3.05
Clinical Competence Average	70.2	3.06	FLEX Weighted Average	68.0	3.51

TABLE 6: Mean scores of students in the Review Program on their most recent FLEX examination prior to entering the course.

b. Letter of Reference

In the application package, candidates were given a form to be completed by their supervisor. This form contained two appraisal items. One asked the respondent to check the sciences which he felt were the candidates particular strengths (Table 7), and the second was a peer ranking of this

person on a six-point scale (Table 8).

With regard to these evaluations, the different total numbers were the result primarily of individuals who had no immediate access to a physician as supervisor or where the supervisor indicated insufficient contact with the candidate to fairly aknowledge his or her strengths and/or weaknesses.

As would be expected, clinical areas were more often judged as strengths in candidates by clinical supervisors, and the class ranking was skewed toward the top end of the scale. The areas identified as strengths most often were psychiatry (34 times) and internal medicine (28 times). The frequency of indicating these two areas as strengths is markedly distinct from all others, which range in frequency from 1 to 10.

c. Self-Appraisal

From Table 9 it is clear that the general trend overall was for the students to rate themselves on the low end of the seven point scale. Rarely did individuals give themselves ratings in the superior range (mean scores of 5 or more) in any science area (Table 10). Therefore as a group, the students indicated that they felt an overall need for knowledge and experience which was greater than their peers. This sense of deficiency in knowledge and experience was general rather than selective for the group. Consequently, those persons who evaluated themselves as superior in some area were clearly distinctive from the class as a whole.

d. Diagnostic Examination

The Diagnostic Examination consisted of two parts administered in two sessions corresponding to the six basic sciences and six clinical sciences, respectively.

In regard to the speed of performance on this examination, Table 11 shows the results of testing time for the class under a condition of no explicit time deadline. The mean time per question was above the 50-52 seconds per

		f			f
1.	Anatomy	7	7.	Medicine	28
2.	Physiology	9	8.	Surgery	6
3.	Biochemistry	1	9.	Ob/Gyn	10
4.	Microbiology	3	10.	Ph/Pm	9
5.	Pathology	7	11.	Pediatrics	8
6.	Pharmacology	9	12.	Psychiatry	34
			13.	Neurology	1

TABLE 7: Science areas checked off on the letter of reference forms as representing the participants' particular strengths (f = frequency of checks)

(1)	Top 5%	25
(2)	Next 10%	22
(3)	Next 20%	12
(4)	Next 40%	5
(5)	Next 40%	2
(6)	Bottom 5%	0
		66

TABLE 8: The frequency each peer rank category was checked for the candidates by their supervisors. (Number of candidates = 66).

	Mean	S.D.		Mean	S.D.
Anatomy	2.28	.94	Medicine	2.81	.92
Physiology	2.56	.87	Surgery	2.58	1.03
Biochemistry	1.96	.84	Ob/Gyn	3.03	1.26
Microbiology	2.51	.99	Ph/Pm	2.18	1.19
Pathology	2.71	.87	Pediatrics	2.71	.98
Pharmacology	2.66	.89	Psychiatry	3.27	1.57

TABLE 9: Mean self-ratings for each science area by the students in the Medical Review Program. (n=45)

	1	2	3	4	5	6	7
Anatomy	16	18	6	5	-	-	-
Physiology	9	15	15	5	-	-	-
Biochemistry	22	14	8	1	-	-	-
Microbiology	11	17	13	4	-	-	-
Pathology	7	22	11	5	-	-	-
Pharmacology	8	18	13	6	-	-	-
Medicine	8	16	16	4	1	-	-
Surgery	15	10	15	5	-	-	-
Ob/Gyn	7	13	15	8	-	2	-
Ph/Pm	18	18	5	3	-	-	1
Pediatrics	10	16	12	7	-	-	-
Psychiatry	9	7	17	5	4	2	1

TABLE 10: The frequency of using each step on the seven point rating scale for the 12 science areas of the self-appraisal form.

Number of items	160		155	
Mean total time	148	min.	144	min.
Mean time per question	55	sec.	56	sec.
Faster than 50 seconds per question	13		19	
Slower than 60 seconds per question	9		22	

TABLE 11: Speed of performance on the Diagnostic Examination administered at the beginning of the FLEX Review Program. (n = 68).

Basic Science Test	r	P	Clinical Science Test	r	P
Total	.742	.001	Total	.794	.001
Anatomy	.853	.001	Medicine	.506	.001
Physiology	.354	.004	Surgery	.482	.001
Biochemistry	.196	.11	Ob/Gyn	.224	.067
Microbiology	.445	.001	Ph/Pm	.073	.5
Pathology	.199	.10	Pediatrics	.358	.003
Pharmacology	.279	.02	Psychiatry	.221	.07

TABLE 12: Results of the split-half correlation of the Diagnostic Exam Part I and Part II as estimates of reliability.

question reported by Hubbard for successful completion of FLEX (3). A considerable number of students were markedly slow and were seen as probably having difficulty even completing a sufficient number of items during a FLEX testing to warrant a passing score. Approximately one third of the class were markedly slow at responding to clinical science test items. This was suspected as resulting from the increased reading requirements indicative of these questions.

The reliability of the Diagnostic Examination was estimated by using the split-half technique. Following Hubbard's (3) procedure, the Spearman-Brown prophecy formula for split-half reliability yielded significant and assuring results for the basic science and clinical science sections of the examination (Table 12). Individual science areas, however, showed considerable variability. Within the basic sciences, questions identified with anatomy, physiology, and microbiology showed the most consistency, while biochemistry and pathology questions were disappointingly weak in consistency. In the clinical sciences, questions in medicine, surgery, and pediatrics displayed the most consistency.

As stated earlier, the validity of the Diagnostic Examination can be attributed to its content review by expert committee. This validity was substantiated by correlation analysis with pre-program FLEX results. Previous FLEX results were viewed as criterion measures for this concurrent validity estimate. Table 13 shows the correlation matrix for the 12 science areas of the Diagnostic Examination and pre-program FLEX.

In general, the basic science portion of the Diagnostic Exam showed more consistency with past FLEX results than the clinical science portion. The results of this analysis also suggested an inter-science clustering or overlap of areas.

The internal consistency of the Diagnostic Examination and the past FLEX scores of the students was examined more fully by an intercorrelation matrix of the two tests (Table 14). The Diagnostic Examination sciences of anatomy

Diagnostic Exam Scores	Pre-Program FLEX Scores											
	Anat.	Physiol.	Bioch.	Path.	Micro.	Pharm.	Med.	Surg.	Ob/Gyn	Ph/Pm	Peds.	Psych.
Anatomy	.39	.52	.04	.29	.00	.33	.01	.07	.20	.28	-.06	.28
Physiology	.44	.54	.15	.45	.05	.64	.00	.20	.43	.49	.16	/14
Biochemistry	-.13	.33	.50	.12	.33	.38	.40	.31	.10	-.11	.47	.23
Pathology	.30	.22	-.03	.34	.05	.15	.02	.09	.30	.26	.19	.09
Microbiology	.03	.09	.44	.19	.46	.38	.46	.43	.17	.00	.53	.13
Pharmacology	.27	.14	-.06	.21	-.13	.25	-.10	.02	.07	.09	.15	-.07
Medicine	.14	.36	.30	.41	.17	.40	.26	.35	.37	.08	.38	-.12
Surgery	-.02	.17	.38	.17	.33	.33	.39	.52	.17	-.04	.40	.03
Ob/Gyn	-.11	.11	.27	-.04	.25	.20	.29	.38	.17	.02	.46	.05
Ph/Pm	.13	-.08	.06	-.16	.13	-.10	-.12	-.12	-.26	-.10	-.05	.18
Pediatrics	-.09	.16	.29	.21	.37	.32	.36	.55	.11	.12	.38	-.00
Psychiatry	.26	.38	.00	.17	-.13	.27	-.03	.10	.14	.28	-.05	.26

r	P
> .30	.05
> .39	.01
> .44	.005
> .56	.001

TABLE 13: The correlation matrix for the 12 science areas of the Diagnostic Examination and the pre-program FLEX scores of the students (n=39). Underlined entries on the diagonal are the critical values.

	Anat.	Physiol.	Bioch.	Path.	Micro.	Pharm.	Med.	Surg.	Ob/Gyn	Ph/Pm	Peds.	Psych.
Anatomy		.63	-.09	.73	-.22	.38	-.36	.05	.54	.57	-.17	-.19
Physiology	.35		.10	.69	-.07	.55	-.17	.03	.53	.60	-.05	-.06
Biochemistry	.15	.43		-.14	.82	.20	.78	.43	-.32	-.37	.72	.56
Pathology	.07	.33	.35		-.18	.57	-.35	.20	.73	.71	-.11	-.27
Microbiology	.05	.37	.47	.32		.08	.84	.50	-.34	-.30	.72	.61
Pharmacology	-.05	-.00	.06	.05	-.22		.00	.41	.49	.37	.15	.00
Medicine	.25	.21	.37	.36	.51	-.38		.54	-.38	-.49	.76	.60
Surgery	.05	.09	.33	.26	.01	.85	-.08		.21	-.01	.57	.29
Ob/Gyn	.00	.30	.36	.21	.32	.00	.20	.20		.66	-.16	-.34
Ph/Pm	.22	.17	.04	.17	.22	-.48	.38	-.33	.13		-.32	-.26
Pediatrics	-.04	-.02	.12	,04	-.23	.96	-.39	.87	.05	-.51		.47
Psychiatry	.00	-.00	.02	.04	-.31	.90	-.39	.75	-.04	-.43	.89	

Diagnostic Exam n = 64

r	P
> .25	.05
> .29	.025
> .31	.01
> .35	.005
> .43	.001

FLEX n = 43

r	P
> .31	.05
> .35	.025
> .37	.01
> .43	.005
> .54	.001

TABLE 14: The intercorrelation of scores of the Diagnostic Exam and the past FLEX for the 12 science areas for the students in the Review Program. Past FLEX results above the diagonal; Diagnostic Exam results are below the diagonal.

and psychiatry appear related in the clinical area of the Diagnostic Examination, while ph/pm is distinctly separate.

The data concerning the Diagnostic Examination was interpretted as indicating a moderately good correlation with students' past performance on FLEX. The Diagnostic Exam, therefore, was dealt with as a concommitant, but separate, source of information regarding the students' medical knowledge. Table 15 shows the results of the class on the Diagnostic Examination by science area. By computer analysis each student's scores were converted to standard scores indicating his or her standing in comparison to the class. This allowed for an internal and an external comparison of each person's performance in each science area. Scores could be rank ordered, yeilding relative strengths and weaknesses within the person's range of medical knowledge and would also indicate his relative position within the group who took the examination. A feedback sheet on Diagnostic Examination results is contained in Appendix O.

When reviewing the results of the Diagnostic Exam with each student, the computer print-out was available so that in addition to the standard scores, the raw percent correct for each section could also be examined. This raw score information highlighted four basic "test-taking" difficulties <u>not</u> related to examination content:

(1) Incorrect identification - errors or ommissions of name or number code on the answer sheet
(2) Incorrect marking - too big or too small answer marks interpretted by machine as errors
(3) Incorrect response sequence - confusion or shifts between columns and rows on the answer sheets
(4) Language deficiency - inability to read or understand the test item, question format, or test instructions.

In order to present an overview of the interrelationships between the four sources of information used to appraise the students' medical knolwedge, the intercorrelations of these data were examined (Table 16). Results of each science area are presented separately. Initial observation discloses a close

	Number of Items	Mean	S.D.
Basic Science Total	160	67.00	11.07
Anatomy	34	14.56	2.84
Physiology	32	11.62	4.09
Biochemistry	28	9.60	2.42
Pathology	26	13.24	2.98
Microbiology	21	8.16	2.82
Pharmacology	19	9.82	2.36
Clinical Science Total	155	71.58	11.58
Medicine	38	19.10	4.07
Surgery	39	16.65	4.56
Ob/Gyn	25	11.41	3.12
Ph/Pm	18	9.48	1.87
Pediatrics	21	7.99	2.58
Psychiatry	14	6.96	2.23

TABLE 15: Diagnostic Examination results for the Medical Review Program participants (Basic Science Test n = 68; Clinical Science Test n = 69).

Past FLEX	Supervisor Ranking	Self Appraisal	Diagnostic Exam		Past FLEX	Supervisor Ranking	Self Appraisal	Diagnostic Exam
	Anatomy					Physiology		
	.56	.54	.39	Past FLEX		.49	.48	.54
.58		.90	.08	Supervisor Rating	.29		.93	.22
-.11	.12		.10	Self Appraisal	-.19	.20		.22
.34	.09	-.13		Diagnostic Exam	.25	.12	.33	
	Pathology					Pharmacology		
	Biochemistry					Medicine		
	-.61	-.39	.50	Past FLEX		-.87	-.25	.26
-.74		.33	-.10	Supervisor Rating	-.42		.24	-.02
-.34	.18		-.20	Self Appraisal	-.38	.20		.17
.46	-.14	.01		Diagnostic Exam	.52	.04	-.07	
	Microbiology					Surgery		
	Ob/Gyn					Pediatrics		
	.55	.14	.17	Past FLEX		-.67	-.28	.38
.66		.08	-.12	Supervisor Rating	-.66		.16	.06
-.13	.07		-.07	Self Appraisal	.11	.16		-.09
-.10	-.11	.02		Diagnostic Exam	.26	.12	.51	
	Ph/Pm					Psychiatry		

TABLE 16: The intercorrelation of four information sources pertaining to the appraisal of students' medical knowledge for each science area.

tie between supervisor ranking and student self-appraisal. Thus, the letters of reference do not appear to add any new information about the students' past FLEX performance. A second notable result of this analysis is the negative relationship between self-appraisal and past FLEX scores.

2. Appraisal of English Proficiency

As indicated in Section IIIC, students were administered the Michigan Test of English Language Proficiency (ELP) and the Michigan Test of Aural Comprehension (AC) as offered by the English Language Institute of the University of Michigan during the three day orientation period. Overall class means were 76.7 on the ELP and 78.9 on the AC. The number of students falling into each of the three classification categories was as follows:

Passed with no further need	38
Highly recommended	8
Definite Need	23

D. Instructional Activity

The challenge involved in attempting to provide a quality educational experience to a group of professional adults with many diverse educational backgrounds wes great.

An interesting observation regarding instructional activity was the hesitation on the students' part to miss any lecture. Exceptions indeed occurred, but basically for reasons of specific lecturer dissatisfaction. This action, on the part of the students, ran contrary to the philosophy of diagnostic appraisal, i.e., to identify areas of strength and weakness, so that each student could select lectures and review topics which required their greater concentration. This observation, however, is in agreement with the students' own self-appraisal, reported on above. The fact that this body of physicians rated their overall knowledge and experience as being less than their peers, explains their hesitation

to miss an important bit of information which might appear on the FLEX."

The didactic portion of instruction was carried out by eighty individuals. This situation proved to be unfavorable in that a lack of cohesiveness or flow was often apparent. Even though the Daily Lecture Schedule differentiated the topics for discussion adequately, there was often the tendancy on the speakers part to cover information already given, or more in line with his specialty. As a result, many topics were not adequately discussed.

The majority of the faculty supplemented their lectures with printed handouts and/or audio-visual materials. Handouts proved to be most popular with the students as they could review the material individually and independent of the lecture. The faculty found this medium helpful, since it allowed them to cover any particular subject with greater speed knowing that the supplement provided additional information. Appendix ? contains the volume of handouts distributed through the program.

Audio-visual materials were particularly popular with clinical faculty. In the absense of a clinical setting, slides provided a suitable alternative to patient contact.

The central role that was envisioned for the Audio-Visual Task Force in technical assistance in software development and investigation did not materialize. Again, this may have been due to the vast number of lecturers. Most indicated that they preferred to use their own materials which they had assembled previously or obtained from a commercial source. These materials did prove to be of excellent quality and met with strong student approval. However, cataloging them into a bibliography for attachment to this final report was not feasible.

Self-instructional materials were made available to supplement deficient areas of the Integrated Curriculum. The series on diagnostic radiology was mentioned earlier in this report. Student response to this material was very favorable, with many students spending hour after hour in its use.

The plans to conduct "Mock Examinations" during the course of this program

also ran into difficulty when it came down to implementation. Little cooperation was displayed by lecturers in supplying multiple choice questions relating to the material they presented. At times, it became awkward to insist upon exam questions from a person whose total lecture committment was only two hours and who had already spent several hours in lecture preparation.

Towards the completion of the program, a series of texts became available which presented questions (and answers) in the formats represented on Days I and II of the FLEX (6). Students were encouraged to purchase these materials and review them individually, not so much for content specifically, but for item format familiarization.

E. Student Affairs

A great deal of time and effort was spent in planning and attempting to anticipate what the non-academic needs and reactions of the program participants would be. The issues of student adjustment to the program living arrangements and anticipated social and psychological difficulties were addressed and methods of intervening in any potential crisis were discussed. It is important to emphasize that in light of the planning that occurred, the severity of some of the problems was totally unpredicted and warranted viewing as moderate to marked adjustment reactions to crises having strong implications as factors effecting ultimate performance on the FLEX. The observations that follow were gathered by the program staff from their daily impressions and also from the personal interview sessions with the clinical psychologist.

Program staff were identified at the outset of the program as being the focal point for student contact regarding adjustment difficulties. A completely unpredicted behavioral pattern was the oppressive and almost total dependency upon program staff for ones personal well being. Students freely communicated

their difficulities and expectations to program staff. The predominant problems arising during the early weeks of the program involved: 1) room arrangements in the dormitory and 2) physical complaints associated with stress. Although room arrangements were initially planned to facilitate the use of English and to provide a culturally enriching experience, many participants were insistent upon sharing rooms with friends or someone of the same nationality to the point that they were willing to accept a refund and withdraw from the program. This necessitated a major and complicated rearrangement in living accommodations.

Early stages of the program were also characterized by many complaints of sleep and eating disturbances, gastrointestinal difficulties, and a general sense of reorientation confusion. Though these occurrences are not unusual for persons traveling or temporarily living in a strange environment, their frequency and the emphasis placed on them by the participants increased their seriousness to a visible level.

As the program progressed into the middle stages of activity, a noticeable shift in problems occurred and began requiring more direct intervention on the part of staff. A number of instances of what appeared to be frank paranoid reactions occurred and were manifested by complaints such as others "intentionally attempting to disrupt my studying so that I fail". In addition, there were many complaints of hoarding and stealing materials, physical threats, and even one individual who proclaimed to have been homosexually accosted in a shower.

Depressive reactions occurred more frequently as the program progressed. It was not uncommon to find program participants waiting for program staff as they arrived for work in the morning indicating that they were unable to concentrate and were falling further behind with the material. Occassional incidences of explosive and angry outbursts by students occurred with one man throwing his tray of food and dishes across the cafeteria counter when not getting what he wanted. Many students were heard in verbal disagreements with one another. Only after

the program ended did one student write back stating that he observed heavy alcohol and substance abuse by a few people in the program.

In addition to these individual problems, there were certain major concerns that affected the class as a whole and generated substantial repercussions.

An event which was external to the program yet produced a major impact upon the course activities was the publication of an article pertaining to the dismissal of state employed physicians in Illinois (Appendix Q). Fifty-seven percent of the class were Illinois residents working in various mental health settings. Since most of these participants were not informed prior to the news release, their reactions ranged from shocked disbelief to angry resentment and bitterness toward their administrators for having been betrayed into signing up for a FLEX review program. A series of crises meetings resulted, consequently postponing the State's actions. However, the resulting pressure and tension persisted throughout the remainder of the program and manifested itself by students leaving the program regularly to attend to personal matters at home. It goes without saying that this action had a deleterious effect upon the effectiveness of the program.

Finally, the class almost unanimously shared the feeling that a conspiracy existed between the Federation of State Medical Boards and individual State Medical Practice Boards with regard to manipulation of test scores. The apparent inconsistencies on the part of State Medical Practice Boards as to FLEX eligibility and score evaluations contributed significantly to this paranoid misconception.

Many students actually believed that their test was identified at the time of writing the examination and that because of personal conflicts with Medical Board staff, they were predetermined to be evaluated as failures.

The source of this misconception is twofold; 1) the FLEX is prepared by the Federation of State Medical Boards and scored by the National Board of Medical Examiners, with the scoring formula being a carefully guarded piece of information

and 2) test results are sent by the NBME directly to the State Medical Practice Boards and not to the individual examinee. It is this stage in the process (when State Medical Practice Boards forward scores to the examinees) that the physicians feel manipulations in the results occur due to personal biases.

V. PROGRAM EVALUATION

Two instruments were available for evaluating the impact of the Medical Review Program in preparing its participants for the FLEX examination. The ultimate criteria was naturally the FLEX examination itself, which was taken by the majority of the participants six weeks following the completion of the program. The second criteria was post-course administration of the Diagnostic Examination. The re-administration of the Diagnostic Examination provided an internal measure of the program's effectiveness with the validity of this instrument being measured by correlation studies with the June, 1977, FLEX results.

With the cooperation of both the students and the Federation of State Medical Boards, June, 1977, FLEX scores were obtained for 59 of the 72 program participants. The results for this group are that 16 of the 59 physicians passed the examination with a FLEX weighted average of 75.0 or better, corresponding to a 27.1% pass rate. Of these 59 physicians, scores were available for 50 individuals. Presumably, 12 participants did not take the June FLEX for unknown reasons.

In terms of the impact of the program on the various sub-scores of the FLEX, a comparison of scores between the June, 1977, FLEX and the most recent pre-program FLEX was made. Table 17 shows the class means for all individuals who reported pre-course FLEX scores and the means for the 50 individuals whose FLEX scores were available after the course. The most obvious distinction in this data is the consistency of improvement in the clinical science areas, but not in the basic sciences. The two exceptions to this general trend are that anatomy showed a significant change in score in the basic sciences while medicine showed no

	Pre-Course FLEX (Most Recent)			Post Course FLEX (June, 1977)				
	Mean	S.D.	n	Mean	S.D.	n	t	p (1 tail)
Anatomy	59.84	5.52	43	63.22	6.23	50	2.773	.005
Physiology	58.63	5.39	43	58.37	6.26	50	<1	N.S.
Biochemistry	57.30	5.28	43	55.42	6.14	50	1.593	N.S.
Pathology	66.14	5.42	43	66.82	7.31	50	<1	N.S.
Microbiology	61.98	4.52	43	61.82	5.49	50	<1	N.S.
Pharmacology	65.56	5.63	43	65.57	6.04	50	<1	N.S.
Basic Science Average	61.62	4.07	43	61.94	5.04	50	<1	N.S.
Medicine	69.39	4.95	44	69.96	5.54	50	<1	N.S.
Surgery	68.91	5.32	44	71.30	6.28	50	1.997	.025
Ob/Gyn	69.57	4.83	44	73.42	6.09	50	3.718	.001
Ph/Pm	66.23	4.48	44	70.65	6.12	50	4.026	.001
Pediatrics	69.61	5.57	44	72.88	4.70	50	3.492	.001
Psychiatry	65.18	8.91	44	69.59	7.25	50	2.625	.01
Clinical Science Average	68.15	3.71	44	71.40	4.59	50	3.795	.001
Clinical Competence Average	70.21	3.06	44	73.08	5.40	50	3.210	.001
FLEX Weighted Average	67.96	3.41	44	70.83	4.95	50	3.291	.001

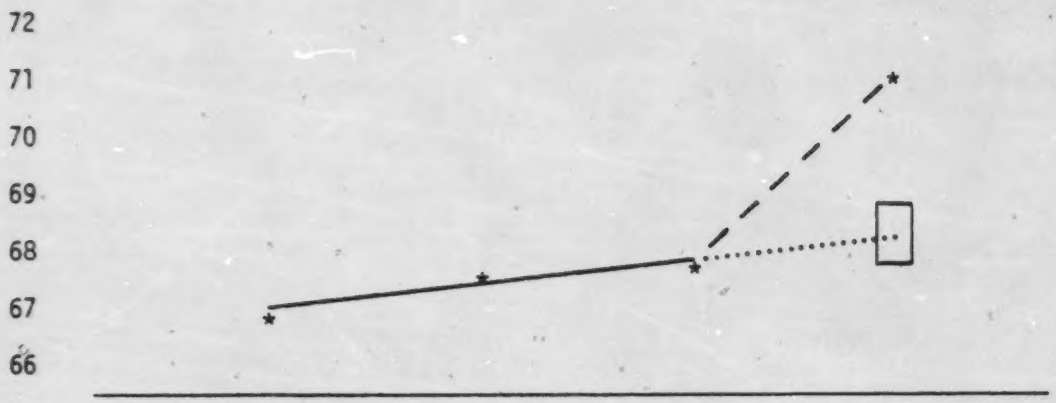
TABLE 17: The comparison of means between FLEX scores of students before and after the Medical Review Program

significant improvement in the clinical sciences.

Another manner in which the FLEX results were interpretted was to analyze the trend of the participants in this program on successive FLEX examinations culminating with the June, 1977, results. By averaging the scores of the participants' most recent pre-course FLEX, their next most recent exam, and the most recent exam before that, a trend could be visualized for the average participant. As would be expected, the number of persons steadily decreases as the number of attempts increases (i.e., fewer people have taken FLEX four times than have taken FLEX three times, and so on). Since the number of students used in the analysis becomes very small, and a number of other factors, such as age, could not be discarded as random or controlled, this material was viewed only as suggestive. Figure 3 shows graphically the trend and most likely prediction that could be made for how the class would perform on the June, 1977, FLEX with no systematic pre-test event. The upward deflection of the FLEX weighted average on the June, 1977, examination can be taken to indicate a positive program effect.

An evaluation of the pre and post Diagnostic Examination results is provided in Table 18. As with the FLEX results, the trend is that the clinical sciences show significant positive gains while the basic sciences do not show consistent improvement. In contrast to the pre-post FLEX results, the sciences of biochemistry and microbiology show improvement, while public health does not show any significant difference.

An interesting analysis conducted with the Diagnostic Examination was a pre-post test comparison of participants' performance as a function of question type. For the three question item formats used in the examination (Type A, one best response; Type B, matching; Type K, multiple true-false), the number of correct responses, regardless of item content, was compared. Table 19 gives the results of this analysis for Part I (Basic Science) and Part II (Clinical Science) of the Diagnostic Examination. Both parts of the test, regardless of content area



	Pre-course Exam III	Pre-course Exam II	Pre-course Exam I	June 1977
Means	66.99	67.58	67.96	70.83
n	10	21	44	50

Figure 3: The trend of scores of persons taking FLEX more than once prior to the Review course. Dotted line indicates most likely predication into the target area. Broken line indicates actual June, 1977, results.

	Pre-Test Diagnostic Examination			Post-Test Diagnostic Examination				
	Mean	S.D.	n	Mean	S.D.	n	t	P
Anatomy	14.56	2.84	68	15.21	3.94	53	1.013	N.S.
Physiology	11.62	4.09	68	12.47	3.89	53	1.166	N.S.
Biochemistry	9.60	2.42	68	10.64	3.00	53	2.06	.025
Pathology	13.24	2.98	68	13.96	3.64	53		N.S.
Microbiology	8.16	2.82	68	10.08	2.63	53	3.86	.001
Pharmacology	9.82	2.36	68	9.89	2.87	53		N.S.
Basic Science Average	67.00	11.07	68	72.25	14.64	53	2.17	.025
Medicine	19.10	4.07	69	21.63	6.36	49	2.33	.01
Surgery	16.65	4.56	69	19.04	6.29	49	2.27	.025
Ob/Gyn	11.41	3.12	69	13.12	3.40	49	2.79	.005
Ph/Pm	9.48	1.87	69	10.20	2.86	49	1.54	N.S.
Pediatrics	7.99	2.58	69	9.55	3.06	49	2.91	.001
Psychiatry	6.96	2.23	69	8.92	2.70	49	4.17	.001
Clinical Science Average	71.58	11.58	69	82.47	20.5	49	3.36	.001

TABLE 18: The comparison of means between the Pre and Post Test administered to students in the FLEX Review Program, Oakland University, February, 1977.

	Pre-Test			Post-Test				
	Mean	S.D.	n	Mean	S.D.	n	t	P
PART I (Basic Science)								
Type A: One Best Response	34.96	6.76	68	39.40	7.49	53	3.31	.001
Type B: Matching	25.10	4.54	68	25.98	8.06	53	<1	N.S.
Type K: Multiple True/False	7.22	2.58	68	6.87	2.98	53	<1	N.S.
PART II (Clinical Science)								
Type A: One Best Response	52.65	8.65	69	62.18	14.66	49	4.07	.001
Type B: Matching	9.10	3.14	69	10.14	3.85	49	1.56	N.S.
Type K: Multiple True/False	9.10	3.14	69	10.69	4.09	49	1.47	N.S.

TABLE 19: Comparison of Pre and Post Diagnostic Examination performance as a function of question item format.

show that the principal improvement occurred with the Type A questions and that students showed no improvement in their ability to handle other more complex question formats. By integrating these findings with the previous results, it appears that while students did make informational gains during the course as reflected on both the diagnostic test and the FLEX scores, the gains were restricted by the students' inability to demonstrate their knowledge in various test item formats.

A method of program evaluation that proved of extreme interest on its own, but also served as a link between the validity of the Diagnostic Examination and FLEX, was a detailed comparison of those individuals who passed the FLEX with those who showed the least improvement on the FLEX. For simplicity sake, these two groups were referred to as the "Best" group and the "Worst" group.

As can be seen in Table 20, the first obvious distinction between the "Best" and "Worst" groups is the age. While more than half of those that passed were less than 45 years old, almost three-quarters of the "Worst" group were in the 46 to 60 year old bracket.* The years since graduation from medical school was an obvious correlate of age ($r = .845$). However, among these two groups, the number of years in the United States was more independent from the persons age ($r = .54$) and, therefore, was considered as separate information. The results suggest that newly immigrated individuals (less than 6½ years in the United States) perform better on FLEX. When the students' past test history is examined, differences appear to exist between the best and worst FLEX performances. The "Worst" group show a history of having difficulty with tests (i.e., 50% took ECFMG and/or FLEX more than once before). In the "Best" group, 73% of those who took ECFMG passed it on the first attempt and close to half of the passing group (46%) had not yet taken FLEX.

* The number of passers used in this analysis was 13 since three persons identified as passing did not have scores reported to us. The "Worst" group was composed of those 15 participants that scored the lowest FLEX weighted average scores on the June, 1977, Exam.

With regard to post graduate training (internships and residencies) no real differences were noted between the "Best" and "Worst" groups. With regard to the type of residency, it appears that those individuals who have had specialty training in Internal Medicine or have taken multiple residencies and completed them performed better on FLEX. The high percentage of psychiatric residencies associated with the "Worst" group may be due to the unusually high number of physicians in the class who were currently working within psychiatric settings due to license restrictions and factors noted prompting the initiation of the review program initially.

Upon examination of the "pre-program" FLEX scores for the "Best" and "Worst" groups (Table 21), a significant distinction is apparent regarding the overall performance of the two groups, while the pattern of scores is remarkably similar (Figures 4 and 5).

The average performance change for each group on the June, 1977 FLEX in comparison to the pre-program FLEX scores is presented in Table 22. Within the basic sciences, the "Best" group demonstrated an improvement of scores while the "Worst" group continued to lose points despite the review program. Exceptions to the general trend are anatomy (both groups showing substantial point gains) and physiology and biochemistry (the "Best" group showed a net loss of points). Within the clinical sciences, the "Best" and "Worst" groups showed discrepant results (Figure 6). The most striking irregularities occurred in public health/ preventive medicine, pediatrics, and the Day III - Clinical Competence section. While the "Best" group showed marked increases in ph/pm and the Clinical Competence Exam, the "Worst" group showed greater point gains in pediatrics than the "Best" group. Another important aspect of this data is that while the "Best" group showed consistent gains in all areas, the "Worst" group showed an average loss in points in medicine and on Part III of the FLEX.

Factor	Best Group	Worst Group
n	13	15
Mean Age	42.54	47.67
% Between 30-45 years old	54 % (7)	27 % (4)
% Between 46-60 years old	46 % (6)	73 % (11)
Sex % Male	85 % (11/13)	53 % (8/15)
Mean years since graduation	15.69	19.53
Mean years in the United States	6.50	14.40
% took ECFMG	85 % (11/13)	80 % (12/15)
% pass first time	73 % (8/11)	50 % (6/12)
% took FLEX before	54 % (7/13)	80 % (12/15)
% more than one attempt	43 % (3/7)	50 % (6/12)
Internship	85 % (11/13)	93 % (14/15)
Residency	69 % (9/13)	53 % (8/15)
Type of Residency:		
Psychiatry	11 % (1/9)	37.5 % (3/8)
Ob/Gyn	22 % (2/9 one person had 2 res. Ob & Psych.)	25 % (2/8)
Internal Medicine	44 % (4/9)	
Surgery	11 % (1/9)	12.5 % (1/8)
Neuro Surgery	11 % (1/9)	
Radiology		12.5 % (1/8)
Public Health	11 % (1/9 one person had surg. & PH)	
Pathology	11 % (1/9)	

TABLE 20: Comparison of characteristics between the BEST (n=13 passers) and the WORST (n=15 lowest) students in Medical Review Program, February, 1977

Area	Best	Difference	Worst
Basic Science Mean	66.39	8.47	57.92
Clinical Science Mean	71.53	7.53	64.00
Clinical Competence Mean	73.34	5.24	68.10
FLEX Weighted Mean	71.37	6.17	65.20
Anatomy	64.86	8.61	56.42
Physiology	62.71	7.46	55.25
Biochemistry	62.86	9.36	53.50
Pathology	72.57	10.90	61.67
Microbiology	65.0	5.83	59.17
Pharmacology	70.43	9.01	61.42
Medicine	75.71	9.29	66.42
Surgery	73.86	8.53	65.33
Ob/Gyn	72.71	7.04	65.67
Ph/Pm	66.71	3.21	63.50
Pediatrics	75.14	10.39	64.75
Psychiatry	65.14	6.14	59.00

TABLE 21: Average Pre-Program FLEX scores for persons in the "Best" and "Worst" groups.

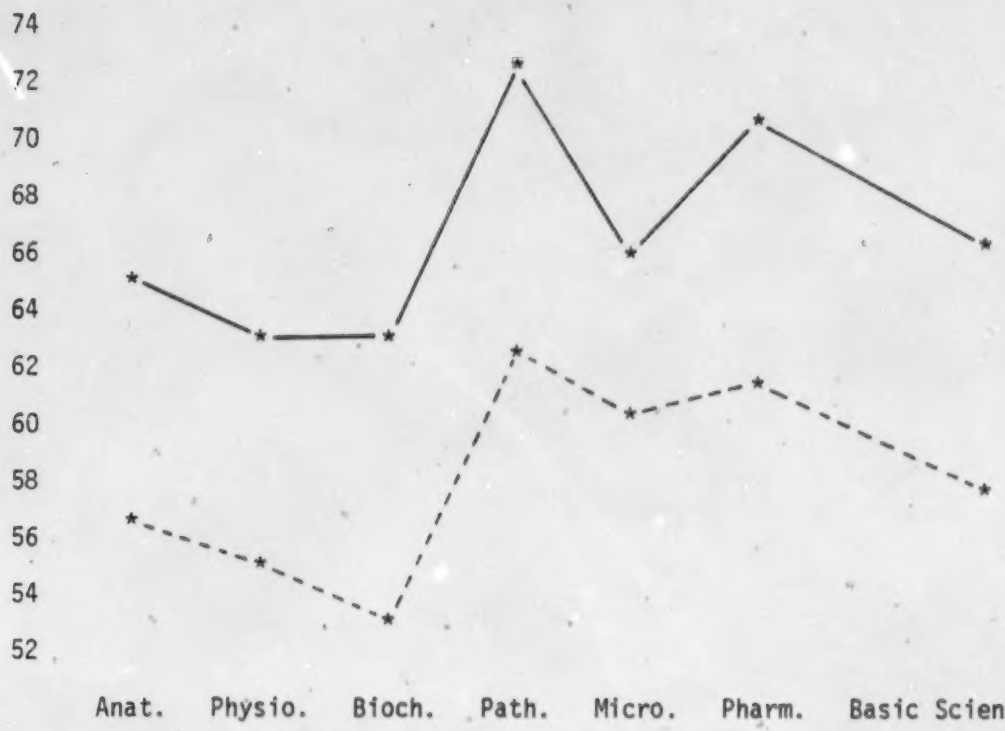


Figure 4: Average Basic Science score for Best and Worst groups in Medical Review Program on their most recent pre-program FLEX attempt ("Best" group *————*; "Worst" group, *--------*)

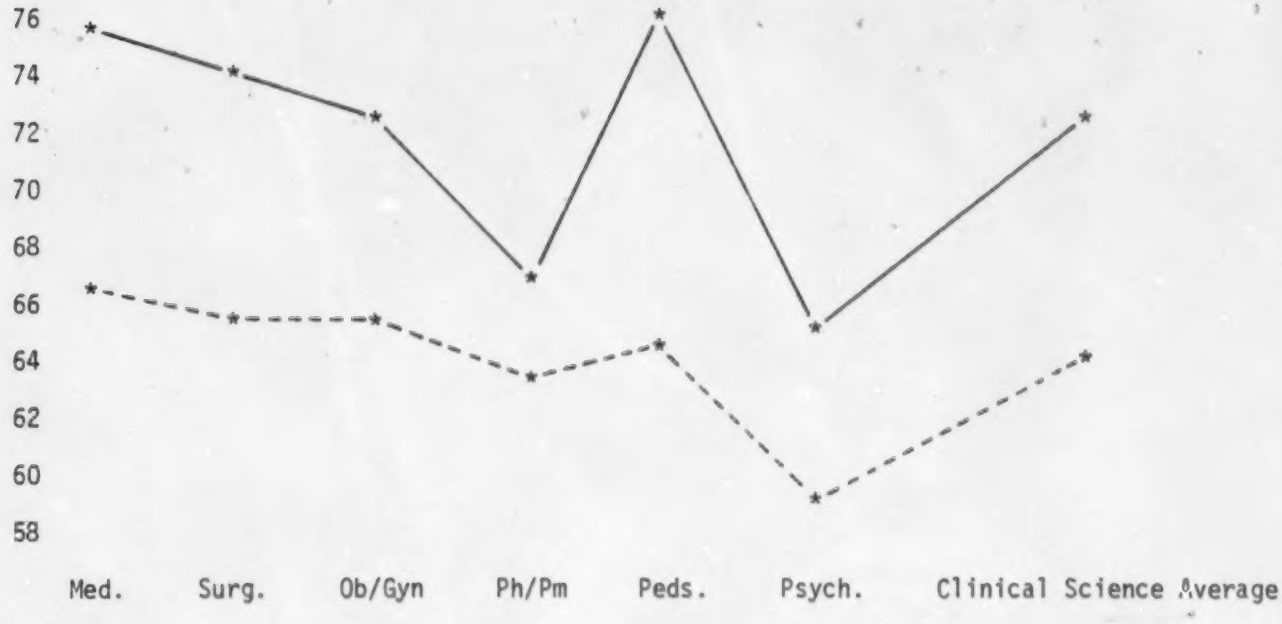


Figure 5: Average Clinical Science scores for Best and Worst groups in Medical Review Program on their most recent pre-program FLEX attempt ("Best" group, *————*; "Worst" group, *--------*)

	Best Group	Worst Group
Anatomy	+ 1.50	+ 2.92
Physiology	- 1.33	- 3.50
Biochemistry	- 1.83	- 3.58
Pathology	+ .50	- .08
Microbiology	+ .17	- .83
Pharmacology	+ 3.33	- 1.17
Basic Science Mean	+ .54	- .85
Medicine	+ .50	- 1.25
Surgery	+ 1.83	+ 1.00
Ob/Gyn	+ 4.33	+ 2.58
Ph/Pm	+ 9.33	+ 2.00
Pediatrics	+ 2.00	+ 4.83
Psychiatry	+ 7.50	+ 5.58
Clinical Science Mean	+ 4.77	+ 2.54
Clinical Competence Mean	+ 6.56	- .48
FLEX Weighted Mean	+ 4.94	+ 0.32

TABLE 22: Average change in points for the "Best" and "Worst" groups on the June, 1977, FLEX in comparison to the pre-program FLEX scores.

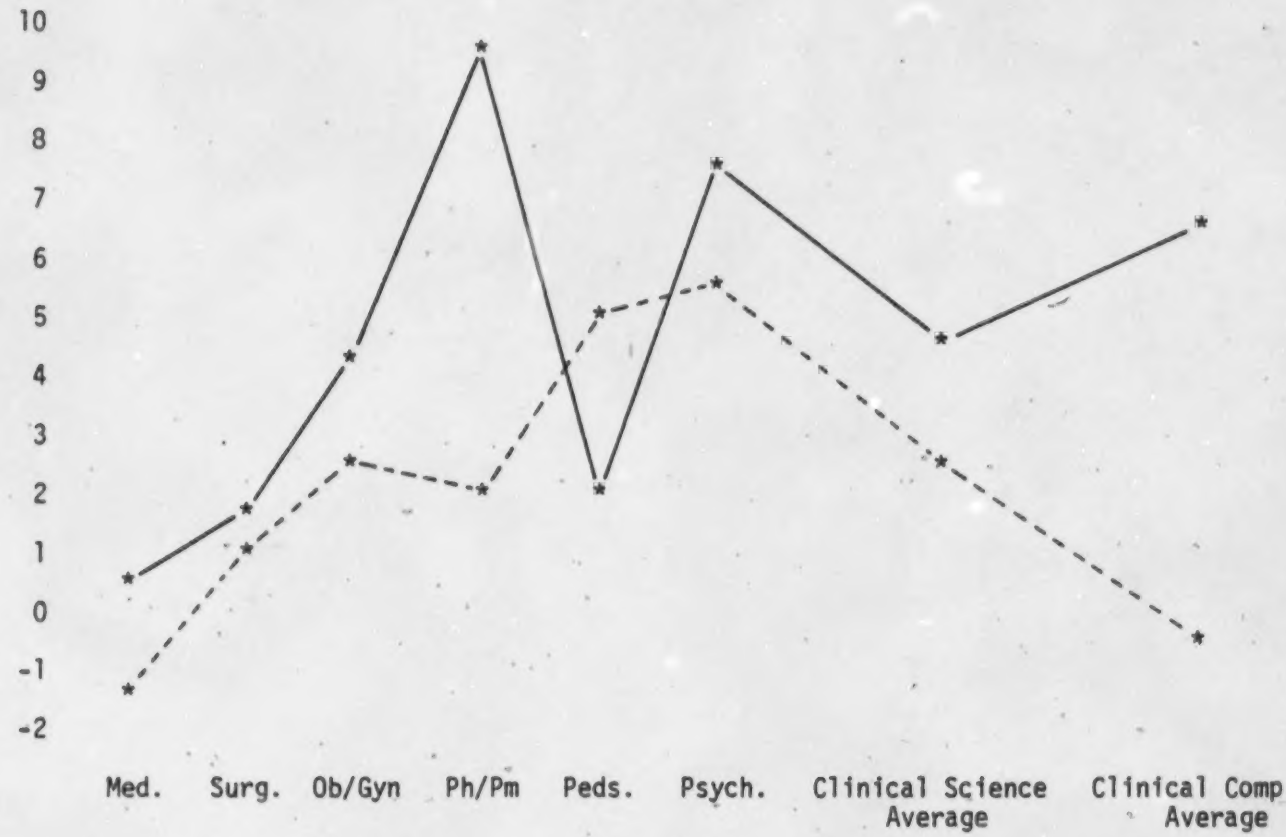


Figure 6: Average change in points for the "Best" and "Worst" groups on the clinical science and clinical competence sections of Pre and Post program FLEX examinations ("Best" group, *————*; "Worst" group, *--------*).

A means of evaluating the Diagnostic Examination as a predictive tool for future FLEX success was accomplished with the use of the "Best/Worst" data established above. Initially, the mean scores for each science area on the Diagnostic Examination were calculated for the "Best" and "Worst" group members. Table 23 gives a comparison of these results. Seven of the twelve science areas show group differences in excess of a full standard deviation, with all areas except psychiatry in the expected direction.

Using the total class average on each science area of the Diagnostic Examination as a cut-off point for classifying a predicted FLEX success or failure, 70.6% of the "Best" students and 64.0% of the "Worst" students were correctly classified (Table 24). Within the basic sciences, physiology and biochemistry correctly classified 87% and 79%, respectively. Pharmacology showed the weakest predictive ability with a 61% correct average. The clinical science questions were moderately successful at classifying the "Best" and "Worst" groups. Surgery gave the greatest predictive ability with 74.5% correct. The psychiatry section was a dismal failure at prediction, even at a chance level. Again, this could be accounted for by the disproportionate number of psychiatrists in the program.

Despite the acknowledged significance of the English/Cultural curriculum to the total program, a means of evaluating its impact upon final examination success was not possible. The only information relating FLEX success or failure with English proficiency and aural comprehension skills, comes by way of further analysis of the "Best" and "Worst" groups. Table 25 shows that the mean scores of the "Best" group on the English Language Proficiency and Aural Comprehension examinations was above the class average, while the "Worst" group fell below class average of both exams. Also, while members of the "Best" group performed consistently on the two exams, the "Worst" group displayed a marked inconsistency (e.g., a score of 70 on the ELP examination and 55 on the AC exam).

	Best Means		Worst Means
Anatomy	15.08		13.73
Physiology	13.30	Significant Difference	9.80
Biochemistry	11.23	Significant Difference	8.73
Pathology	15.00	Significant Difference	11.87
Microbiology	9.23	Significant Difference	6.67
Pharmacology	10.00		8.87
Medicine	21.23	Significant Difference	17.20
Surgery	19.38	Significant Difference	14.87
Ob/Gyn	12.31		10.87
Ph/Pm	9.60		9.27
Pediatrics	8.54	Significant Difference	6.87
Psychiatry	6.85		6.87

TABLE 23: Comparison of the Diagnostic Exam Results for the Best and Worst groups. The S.D. for the whole class (n=69) is used as an approximation for significant descrepancy.

		Number correctly identified for each group and percent correct	
		Number Above Mean	Number Below Mean
Science	Mean	Best Group (n=13)	Worst Group (n=15)
Anatomy	14.55	9 (69%)	10 (66%)
Physiology	11.68	12 (93%)	12 (80%)
Biochemistry	9.75	11 (85%)	11 (73%)
Pathology	13.18	9 (69%)	11 (73%)
Microbiology	8.29	9 (69%)	10 (66%)
Pharmacology	9.65	8 (62%)	9 (60%)
Medicine	19.19	9 (69%)	10 (66%)
Surgery	16.71	9 (69%)	12 (80%)
Ob/Gyn	11.49	9 (69%)	9 (60%)
Ph/Pm	9.69	8 (62%)	9 (60%)
Pediatrics	7.97	10 (77%)	10 (66%)
Psychiatry	6.86	7 (54%)	3 (20%)

TABLE 24: Classification of Best and Worst students using the class means in each science area on the Diagnostic Exam as a cutting score.

English Competency Examination

	"Best" Means	Class Means	"Worst" Means
ELP Exam	81.36	76.71	75.00
AC Exam	81.18	78.90	76.67
Classification by scores:			
I (85 - 100)	6		3
II (65 - 84)	2		6
III (64 or less)	3		6
Consistency of 2 scores by category	8/11 (73%)		5/15 (33%)

TABLE 25: Performance of the "Best" and "Worst" groups on the English Language Proficiency (ELP) and Aural Comprehension (AC) examinations in comparison to the entire class average.

VI. DISCUSSION

Statistical information gathered from the Federation of State Medical Boards allows an external measure of the success of the Medical Review Program and indicates that the overall pass rate for Foreign Medical Graduates on the FLEX is just under 50%. As reported in the previous section of this report, the overall pass rate on the June, 1977, FLEX for participants in the Medical Review Program was 27.1%.

A closer analysis of the results of the 59 physician participants (out of a class of 72) whose results are known, indicates that 17 physicians were taking the FLEX for the first time with 7 subsequently passing for a pass rate of 41.2%. This compares with an anticipated pass rate of approximately 60% according to the Federation for first-time FMG examinees (2). Of the remaining 42 participants (who were repeating the FLEX), 9 passed for a 21.4% pass rate. This compares to an approximate 25% pass rate for FMG repeaters according to the Federation (2).

Initial inspection of the above data suggests that the Medical Review Program did not contribute to improving the pass rates of the participants on the June, 1977, FLEX and in fact, produced a negative effect even though the evaluation section of this report provides a detailed statistical analysis of the program participants' scores and demonstrates significant score improvement in the clinical competence and clinical science areas with less improvement in the basic sciences.

The "Best/Worst" group analysis conducted in the evaluation section of this report suggests that certain characteristics such as age, number of years in the U.S., and number of years since graduation from medical school, may influence an examinees liklihood of success on the FLEX even following participation in a comprehensive medical review program.

The contract admission criteria clearly distinguished two groups of participants who perhaps by chance represented extremes in the above characteristics. The non-fully licensed yet fully-employed (NFLFE) group of participants (48 of the 59 persons whose FLEX results were known) presented an average age of 48 years, had been out of medical school 19.5 years, and had been in the U.S. an average of 16.0 years. The bona fide refugee (BFR) group (n=11; 9 Vietnamese, 2 Russians) presented an average age of 36 years, had been out of medical school 9.5 years and had been in the U.S. 1.8 years.

Regarding the June FLEX success of these two groups, the refugee group had a pass rate of 63.6%, which is approximately 15% above the national norm for FMGs, and the non-fully licensed yet fully-employed group had a pass rate of 16%, which is 31% below the anticipated national norm for FMGs.

In attempting to understand why a 46% difference in the pass rates occurred between the two groups, two sources of information were analyzed.

The first and most obvious criteria would be a deficiency in English language proficiency and lack of cultural orientation. Analysis of the BFR and NFLFE groups indicated that test results on the English Language Proficiency and Aural Comprehension examinations yielded mean scores of 80/82 and 79/81, respectively, for no significant difference in English language proficiency. By contrast, however, the NFLFE group had been in the U.S. an average of 19.5 years versus 1.8 years for the refugee group, indicating no correlation between cultural orientation and FLEX success.

The second factor warranting investigation was the level of familiarity and exposure to the multiple-choice objective format of the FLEX examination. Within the refugee group, 90.9% had not previously taken the FLEX versus 14.6% of the NFLFE group. This data suggests that familiarity and exposure to the multiple choice objective format of the FLEX does not predispose one to success.

It is fully recognized that the above data is not in any way conclusive and that even though a whole host of arguments relative to nationality, motivation, medical education system, etc., can be advanced to explain the significant differences in pass rates between the two groups, it is the hypothesis of the authors of this report that objective, quantifiable characteristics such as those identified in the "Best/Worst" group analysis of this report can be projected to predict levels of success on the FLEX following participation in a Medical Review Program such as outlined in this report.

To carry this analysis one step further, a scale was developed (see Appendix R) assigning positive points to those characteristics indicative of the "Best" group while negative points were granted for "Worst" group features.

This analysis was then applied to 23 participants in an identical offering of the Medical Review Program conducted at Oakland University in September, 1977, on a self-supporting basis.

The application review process yielded a net positive score for twelve participants (predicted passers) while eleven individuals (predicted failers) received net negative scores.

Examination results for the December, 1977, FLEX indicate that six of the twelve predicted "Passers" did indeed pass, for a 50% predictive success. None of the eleven predicted failers passed, for a predictive success of 100%.

Based upon this limited trial, it appears less difficult to identify participants for whom the Review Program offers limited opportunity to succeed. Justification, for the possible use of an Application Review System scale occurs as a result of the fact that the "Best" group or predicted "Passers" demonstrated point gains in all areas of the FLEX and an improved FLEX weighted average of 4.94 points when compared to pre-program exam results, while on the other hand, the "Worst" group or predicted "Failers" actually demonstrated a net loss of points

on the basic science section of the FLEX and a FLEX weighted average increase of only 0.32 points.

It is beyond the scope of this report to deal with the issues of measuring medical competency, licensing examinations, and licensing regulations. However, while the data base for this study is much too small to provide conclusive evidence, the data does suggest the need for further investigation into the validity and effectiveness of conductive medical review programs of similar design and duration for physicians who characteristically fall into the "Worst" category.

While no one would argue the need for an evaluation of medical competency, the question is whether the current instrument (FLEX) is appropriate for the population it is attempting to measure and for whom this contract was specified.

An interesting correlation study would be to compare results of the non-fully licensed fully employed FMG examinees with fully employed fully licensed USMGs having identical characteristics, on an identical examination. This would be particularly interesting with regard to the basic sciences.

VII. SUMMARY

In July, 1975, the Bureau of Health Resources Development, Department of Health, Education and Welfare, awarded contract no. 231-75-0607 to Oakland University, Rochester, Michigan for the purpose of developing and administering a preparatory course for the Federation Licensing Examination (FLEX) for non-fully licensed physicians employed in the United States.

Statistical data indicated that nearly 75% of the candidates for licensure in states employing the FLEX were foreign medical graduates who had a pass rate just under 50% compared with a pass rate of about 85% for United States medical graduates.

Twenty-one months were spent in program development which produced an integrated basic and clinical science curriculum and an English/cultural curriculum which were presented during a three month (13 week) program. In addition, an elaborate Diagnostic Appraisal process was developed for the purpose of identifying participants' individual strengths and weaknesses prior to program participation.

One hundred fifty-seven applications were received with seventy-two participants appearing for the program in January, 1977. The mean age of the class was 46.7 years, which was older than anticipated. Two-thirds were men and one-third women. The majority had been in the United States an average of 14 years.

The greatest population of the class resided in the Midwest (79%) with the State of Illinois accounting for over half of the participants (58%).

Forty-five different medical schools were represented by the class and the average number of years since graduation from medical school was 48.6 years.

Fifty-nine of the seventy-two participants took the June, 1977, FLEX and performed with a pass rate of 27.1%, which is approximately 22% below the overall anticipated national norm.

A detailed statistical analysis of the program participants' scores on the June, 1977, FLEX demonstrates significant score improvement in the clinical competence and clinical science areas with less improvement in the basic sciences when compared to pre-program FLEX scores.

As a result of the stipulated contract admission criteria, two distinct groups of participants could be identified. Those physicians who were non-fully licensed and yet fully employed presented an average age of 48 years, had been out of medical school 19.5 years, and had been in the United States an average of 16.0 years. This group of physicians had a pass rate of 18% or roughly 31% below the anticipated national norm for FMGs.

By contrast, the bona fide refugee group presented an average age of 36 years, had been out of medical school 9.5 years and had been in the United States 1.8 years. This group had a pass rate of 63.6% which is approximately 15% above the national norm.

Based on the obvious differential in passing rates between the two groups, a more elaborate analysis of personal and demographic data was conducted resulting in an Application Review System which attempts to identify those physicians who can benefit from participation in a similar Medical Review Program. This analysis was subsequently applied to 23 participants in a repeat of the Medical Review Program conducted at Oakland University on a self-supporting basis in September, 1977, resulting in a 50% predictive success for "Passers" and 100% predictive success for "Failers".

Though this data is not conclusive, it does demonstrate the need for further investigation into the validity and effectiveness of conducting future Medical Review Programs of similar design and duration for physicians who present characteristics which appear to predispose them to high liklihood of failure on the FLEX regardless of review program support.

Bibliography

(1) Request For Proposal No. 231-75-0607 (P) for developing and administering a preparatory course for the Federation Licensing Examination (FLEX), Bureau of Health Resources Development, Department of Health, Education, and Welfare, April 10, 1975.

(2) Merchant, Frederick T., General Chairman, FLEX Test Committee, personal communication.

(3) Hubbard, John P., Measuring Medical Education, Lea and Febiger, Philadelphia, 1971.

(4) FLEX Description of Examination pamphlet, 4th Edition, The Federation of State Medical Boards of the United States, Inc., Forth Worth, Texas, 1976.

(5) Nordlund, James J., Ed., Internal Medicine, Patient Management Cases: Pretest Self-Assessment and Review, Pretest Service, Inc., Wallingford, Conn., 1977.

(6) Pretest Self-Assessment and Review Series (Anatomy, Biochemistry, Physiology, Microbiology, Pharmacology, Pathology) Pretest Service, Inc., Distributed by Blakiston Publications, McGraw-Hill Book Company, 1976.